PB-249 891

A BENEFIT/COST EVALUATION OF DRINKING WATER HYGIENE PROGRAMS

J. Edward Singley, et al

Florida University

Prepared for:

Environmental Protection Agency

November 1975

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

BIBLIOGRAPHIC DATA SHEET	1. Report No.	2.	3. Recipient's Accession No.

4. Title and Subtitle
A Benefit/Cost Evaluation of Drinking Water Hygiene Programs

5. Report Date
Nov. 1975

6.

7. Author(s)
Singley, J.E., Hoadley, A.W., and Hudson, H.E. Jr.

8. Performing Organization Rept. No.

9. Performing Organization Name and Address
University of Florida
Gainesville, Florida 32611

10. Project/Task/Work Unit No.

11. Contract/Grant No.
68-01-1838

12. Sponsoring Organization Name and Address
Office of Water Supply
Environmental Protection Agency
Washington, D.C. 20460

13. Type of Report & Period Covered

14.

15. Supplementary Notes

16. Abstracts

A methodology has been developed for the evaluation of the benefit/cost ratio of water hygiene programs. The methodology is based upon regressions models.

This methodology has been used to evaluate the benefit/cost ratio for water quality improvements that effect the incidence of the diseases salmonellosis, shigellosis and hepatitis.

The regression methodology used could be used for other water programs subject to the availability of the appropriate data.

17. Key Words and Document Analysis. 17a. Descriptors

water hygiene
cost/benefit ratios
regression models
salmonellosis
shigellosis
hepatitis

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

PRICES SUBJECT TO CHANGE

18. Availability Statement	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 136
	20. Security Class (This Page) UNCLASSIFIED	22. Price $6.00

FORM NTIS-35 (REV. 10-73) ENDORSED BY ANSI AND UNESCO. THIS FORM MAY BE REPRODUCED USCOMM-DC 8265-P74

063018

PB 249 891

A BENEFIT/COST EVALUATION OF DRINKING WATER HYGIENE PROGRAMS

BY

J. Edward Singley
A. W. Hoadley
H. E. Hudson, Jr.
College of Engineering

Edna T. Loehman
Institute of Food & Agricultural Sciences

University of Florida
Gainesville, Florida 32611

NOVEMBER 1975

Contract No. 68-01-1838

Prepared For
Office of Water Supply
U. S. Environmental Protection Agency
Washington, D. C. 20460

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
US Department of Commerce
Springfield, VA. 22151

ABSTRACT

A methodology has been developed for the evaluation of the benefit/cost ratio of water hygiene programs. The methodology is based upon regression models.

This methodology has been used to evaluate the benefit/cost ratio for water quality improvements that effect the incidence of the diseases, salmonellosis, shigellosis and hepatitis.

The regression methodology used could be used for other water programs subject to the availability of the appropriate data.

CONTENTS

Section		Page
I	Conclusions	1
II	Recommendations	2
III	Introduction	4
IV	Extent of Water Quality Problem: Measuring the Costs of Infectious Disease Associated with Drinking Water	12
V	Methodology for Analysis of Benefits of Drinking Water Improvement Programs	28
VI	Cost of Drinking Water Quality Improvement Program	49
VII	Benefit-Cost Analysis of a Drinking Water Quality Improvement Program	60
VIII	Acknowledgements	67
IX	References	68
X	Publications	76
XI	Appendicies	77

FIGURES

Number		Page
1	Average Annual Waterborne Disease Outbreaks, Reported vs Non-reported, 1946-70	6
2	Cost of Construction of Water Treatment Plants: 1974 Dollars	54
3	Cost of Carbon Adsorption Treatment	56

TABLES

Number		Page
1	Reported Infectious Disease	14
2	Costs of Illness	20
3	Costs of Hospitalization	22
4	Costs of Mortality	23
5	Costs of Generated Income Lost	24
6	Distribution of Costs of Three Infectious Diseases by Cause and Type of System	25
7	Distribution of Costs of Infectious Hepatitis by Cause and Type of System	26
8	Costs of Infectious Disease Associated With Drinking Water	27
9	Infectious Hepatitis	34
10	Shigellosis	34
11	Salmonellosis	35
12	Variables Used in the Regression Analysis	35
13	Comparison of Sample, U.S., and Predicted Values	37
14	Reported Cases Predicted to be Water Related, U.S., 1969	38
15	Characteristics of 870 Water System Sample	40
16	Regression Equation for LSPCD	41
17	Improvement Programs and Their Effects on Average Drinking Water Quality	43
18	Reduction in Average Reported Disease Rates Due to Improvements in Drinking Water Quality	44
19	Stability of Water Distributed by 100 Largest Cities in U.S. (Based on Langelier Index)	57

20 Summary Estimated Costs of Various Water Quality Programs and Estimated Reduction in Water Borne Disease---------- 58

21 Characteristics of Existing Public Water Supply Systems and of Measures to Improve Water Quality---------- 60

22 CWSS Sample Characteristics---------- 61

23 Distribution of 19,007 Municipal Water Facilities by Size of Community Served, 1963---------- 62

24 Estimated System Characteristics for 40,000 Community and Municipal Water Supplies---------- 62

25 Capital Costs for Adding Disinfection and Upgrading Deficiencies in Disinfection 40,000 Systems in the U.S.---------- 63

26 Summary of Additional Costs for the Improvement Program------ 64

27 Benefit/Cost Ratios---------- 66

SECTION I
CONCLUSIONS

1. A method for analyzing the effects of water quality improvement programs, in terms of a benefit/cost ratio, has been developed and demonstrated.

2. This method was used to estimate the ratio for a specific improvement program for illustrative purposes and considered the diseases salmonellosis, shigellosis and infectious hepatitis.

3. The program analyzed was based on the 1970 Community Water Supply Study by the Public Health Service. The factors considered most important as to their effect on water quality, on which data were available, were adequate disinfection, surveillance, presence of a trained operator and presence of a chlorine residual.

4. In addition to the analyzing program effectiveness, the disease equations developed give a way of estimating how much of reported disease is actually related to water, either directly or indirectly. It was shown that water related disease is actually much higher than that reported as water related.

5. The importance of the method is that it allows program-specific comparisons to be made for a meaningful benefit/cost analysis.

6. For the program analyzed, it was estimated that the total costs of the waterborne infectious diseases considered are $156,448,000 per year.

7. The greatest costs associated with waterborne infectious hepatitis can be attributed to public systems using untreated ground waters ($46,150,000 cut of a total of $114,900,000).

8. The purpose of these predictions was to test the validity of the models; the results indicated that the hepatitis and salmonellosis equations were valid as models of reported disease rates.

9. The approach should be useful for evaluating benefit/cost ratios for water hygiene programs, if the appropriate data are or can be made available or if reasonable estimates can be made.

SECTION II
RECOMMENDATIONS

Although a benefit/cost methodology for water quality improvement was successfully demonstrated here, there are some weaknesses in the analysis due to problems with the data on which the analysis is based. Much of the data used here came from county health officials and the Community Water Supply Study (CWSS). The CWSS was a very valuable source but did contain some contradictory information and many gaps. Thus, the following recommendations have mainly to do with improving the quality of the data available for use in benefit and cost measurements. With better data, more accurate benefit/cost estimates can be made.

Although an ideal benefit/cost methodology requires both cost and benefit functions, it was not possible here due to lack of data to develop a cost function relating levels of water quality to minimum cost of achievement. (Instead to obtain a benefit/cost ratio for a specific program, costs of the program were estimated using CWSS data as a basis.) The CWSS data contained little information about operator characteristics, treatment, and cost of improvement, due to the poor response to these questions. Determination of a cost function requires knowledge of what levels of water quality results from various operator, surveillance and treatment combinations as well as their costs. Such relationships are not definitively known although an attempt was made here to develop them. Furthermore, the cost savings which may occur through consolidation of small systems need to be investigated. All these considerations relating to development of a cost function should be further researched to develop a better benefit cost policy tool.

The accuracy of the benefit function developed here is also somewhat limited due to data problems. The measure of water quality from the CWSS which was related here to disease was based on a one time sample of water quality; a measure of water quality based on samples taken over a year would give a clearer picture of water quality in a community. Disease data were limited to that reported to county health authorities; this is an area which would also benefit from better data collection efforts. Placing a value per case of infectious disease involved some difficulties due to the lack of information on factors such as length of illness and cost of medications; such limitations can only be overcome through more involvement of doctors in waterborne disease research. Furthermore, disease data should be collected for diseases which may be related to water quality other than infectious and viral diseases.

Better assessment of benefits and cost of water quality improvement in rural and high risk areas is important due to the large number of water systems in such areas. Since the CWSS data were limited mainly to SMSA's this assessment also requires the collection of data in these areas.

Thus, though a benefit/cost methodology was successfully developed and its usefulness demonstrated here, such analyses could be much improved through more complete surveys of water systems and community health. Such surveys made on a regular basis would also help to identify trends in water quality---disease relations and the adequacy of water supply systems.

SECTION III
INTRODUCTION

Purpose

The purpose of this study was to develop and test a model approach which can be applied to the analysis of costs and benefits associated with the programs and efforts of the Division of Water Supply of EPA dealing with the relation of human health to drinking water quality. The methodology must provide tools which can be applied in the evaluation of the existing and proposed programs of the Division of Water Supply and identification of future problems for the purpose of: 1) establishing appropriate levels of funding for programs; 2) establishing priorities; 3) allocating available funds; 4) stimulating private and public sectors to undertake programs which will be most productive and beneficial; and 5) educating the public and legislative bodies regarding short and long range implications of drinking water quality in relation to health.

Background

There has been considerable discussion in the history of public health regarding the role of impure water in disease and ill-health. Nevertheless, there is a great variety in the precision with which certain cause and effect relationships can be made; only a few direct associations have been clearly identified from daily experience. This is despite the fact that cause and effect can be clearly demonstrated in the laboratory. In addition, there has been little attempt to accurately quantify the benefits that have resulted from programs devoted to the improvement of drinking water quality and the resultant reductions in ill health. A major problem has to do with estimating the number of disease cases due to poor water quality.

Although proposed drinking water standards [8] would cover both bacteriological and chemical contaminants affecting health, this study focuses only on disease related to bacteriological and viral contaminants. The effects of chemical contamination are long term and often difficult to identify with a single disease. The diseases to be studied here are infectious hepatitis (viral), and salmonellosis and shigellosis (bacteriological). Hepatitis may be associated with other causes besides water (e.g., food and drug abuse). Salmonellosis may be foodborne as well as waterborne. Proportionally more shigellosis than salmonellosis may be related to drinking water [7] since shigellae are primarily associated with man in contrast to salmonellae which are strongly associated with animals [7]. These diseases were chosen for study because cases of these diseases must be reported to the Public Health Service so that information on them is available. The method of analysis used here will be applicable to other diseases and water contaminants as more information becomes available.

Even though by law cases of the above diseases must be reported, there is much uncertainty about the actual number of cases which occur as well as about the source of the disease (due to water, food, or person-to-person contact). There are two basic problems. First, regardless of cause, detection of disease depends on the gravity of the symptoms and whether the patient feels sick enough or can afford to see a doctor. The reporting of detected cases then depends on the compliance of local doctors and the efficiency of the State Health Department in collecting the information which in turn depends on the emphasis placed by each state on health programs.

Second, of the reported disease, the source of the disease may go undetermined unless the size of an outbreak is large enough to be recognized. The investigation of outbreaks is often incomplete and conducted long after the outbreak has occurred making causality difficult if not impossible to determine. Epidemiological studies [1,2] have shown that there may in some cases be at least twelve times as many water-caused outbreaks as are reported to the Center for Disease Control (CDC) and that in some outbreaks as few as one percent of the actual cases are reported [7]. Figure 1 from Craun and McCabe [4] shows that the number of unreported cases of waterborne disease may be increasing.

In addition, reporting of disease by cause rarely includes those cases indirectly caused by water, i.e. those secondary cases caused by contact with individuals who contracted disease through water. These problems make it very difficult to estimate benefits of better water quality since it is not known how many cases of disease are actually related to poor water quality.

This study will deal with the problem of estimating the number of water-related disease cases (those caused directly or indirectly by water) by estimating regression equations relating the rate of reported cases of infectious disease to the quality of drinking water and other factors and estimating how changes in water quality will change the number of cases reported. (To determine actual reduction in disease would then require the use of multipliers to adjust reported disease by the proportion of actual to reported disease.) Although regression has previously been used to analyze the effects of air pollution abatement on human health in terms of reduction in mortality [10], the approach used here is unique in its application to the analysis of the benefit of disease control. This study is possible only because of the availability of data on both bacterial water quality from the CWSS and on reported disease obtained directly from county and state health departments for the 39 counties included in the 11 states covered by the CWSS.

Limited data are available for such an analysis with regards to selected quality-health relationships. In this era of increased emphasis on the need to demonstrate the viability, productivity, and

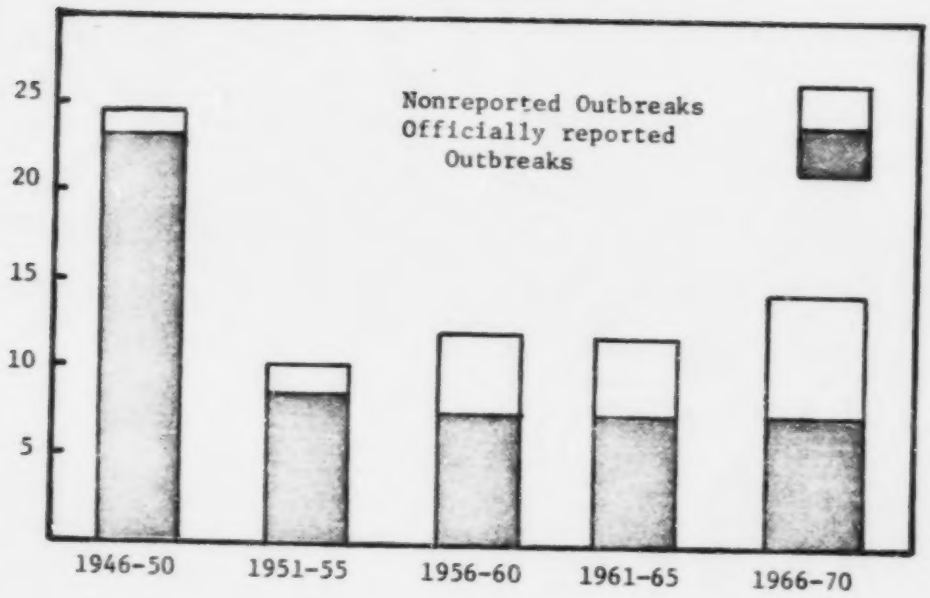


Figure 1. Average Annual Waterborne-Disease Outbreaks, Reported vs. Nonreported, 1946-70.

public benefits derived from government programs, it is necessary to assimilate, analyze and report on the data as it relates to the direction (redirection) of program and/or program components. Other methods must also be developed to aid in evaluating public benefit potential where data are lacking.

During the recently completed Congressional Session the Congress passed and the President signed, a Safe Drinking Water Act. With the new authorities mandated by this legislation, the emphasis will be on related drinking water programs, in addition to the needed capability to isolate high payout programs both within existing authority and the new legislation the results of this study should be very useful in the management of water supply programs.

Since the Environmental Protection Agency will be responsible for developing and recommending national policy regulations and guidelines for authorities concerned with planning, developing, coordinating, and administering water supply programs; it will have to:

1. Develop technical guidance for regional offices to carry out the regulations concerning water supplies serving public carriers;
2. Establish health criteria and recommend standards for the quality of water supply systems;
3. Establish and maintain a national register of public water supply systems; and,
4. Support or conduct studies, training, and technical assistance in the water supply field.

With these requirements which will expand the applicability of drinking water standards beyond interstate carriers, it will be increasingly necessary to expend limited federal dollars for environmental/health programs in those areas where the greatest benefits will occur.

Approach

The task, as it was viewed for the purposes of the study consisted of the following:

1. To propose a methodology which can be applied to the existing and proposed programs of the Division of Water Supply of EPA and satisfy the purpose stated above.
2. To collect historical data required to demonstrate the applicability of the methodology proposed.
3. To analyze the availability or lack of availability, accessibility, and usefulness of data required for testing the methodology.
4. To apply the methodology to the analysis of costs and benefits associated with a model disease and/or a model water quality parameter and to demonstrate its value in

program planning.

5. To recommend procedures for the application of the methodology to the analysis of costs and benefits associated with diseases or water quality parameters the significance of which is not well established.
6. To recommend ways in which the study could be extended to fuily satisfy the over-al purpose as stated at the first of this paper.

Development of Methodology

A preliminary model was developed to guide collection of data. Incorporated in the model were sources of social cost due to disease on cost of prevention, as follows:

1. Social costs---
   a. Determination of the extent of the health problem and health costs (excluding home treatment)
   b. Loss of productivity of those affected and their families
2. Costs of prevention---
   a. Operator training and improvement
   b. Treatment
   c. Alternative sources
   d. Surveillance
   e. Enforcement

Identification of such costs required data on the characteristics of drinking waters throughout the country, methods of water treatment and their costs, incidence of disease throughout the country, the existence of surveillance programs and their costs, costs of medical treatment, use of medical facilities, etc. The incidence of disease had to be considered in terms of its endemic and epidemic occurrence not only in the population at large but in special risks groups. As specific data was obtained the model was modified to accomodate new understanding of the problem.

General Concepts

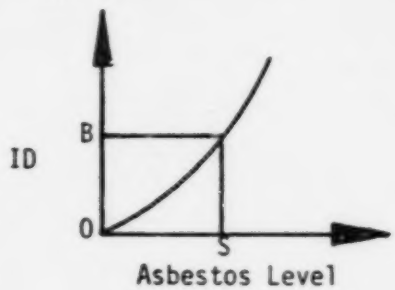
Benefits to implementation and/or improvements of water hygiene programs are related to three categories of diseases: 1) Chronic toxics, such as asbestos in water; 2) Poisons, such as mercury or lead in water; and 3) Infections, due to bacteriological and viral factors in water. Benefits/cost models may have to be developed for each type, both because prediction of benefits may vary in each case and different policy variables may be of concern in each case. However, the general process of analysis in each case is the same and requires the four steps outlined:

1. Determination of the incidence of disease (ID) for the

different categories of waterborne diseases. Since ID's may differ from documented outbreaks of disease attributed to drinking water, problems were faced in: (a) Estimating the actual cases of disease which may not have been reported; and (b) Identifying chronic illnesses due to water, which did not follow the usual epidemic patterns.

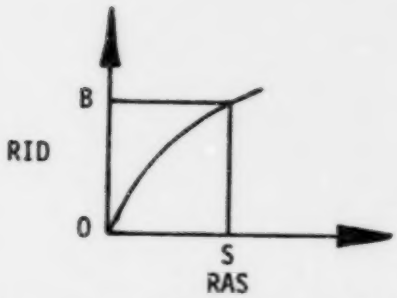
2. Determination of the benefit function. Benefits, i.e., costs averted by a particular program, can be computed by assigning a value to ID. Total benefits (TB) is the product of value (V) and ID. TB = (V)( ). However, merely taking the actual occurrences 1961-1970, and multiplying by V does not give information about possible benefits resulting from alternative programs. Such estimates merely indicate the maximum benefits attainable due to the complete reduction in disease. To be able to analyze benefits of specific programs it is first necessary to determine the functional relationships, between reduction in disease and water quality. This require the following two steps:

   a. Determine ID as a function of a given standard (level of contamination of water with, say asbestos). Graphically the physical relationship between these variables might appear as:



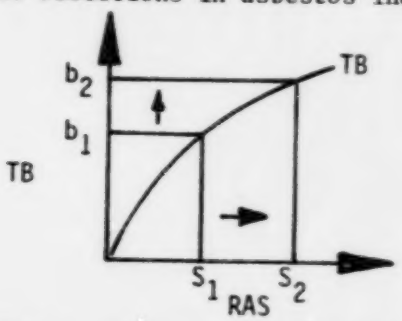
As the amount of contaminant or the asbestos level in water increases the ID is expected to increase. OS might be considered to be the present standard for asbestos levels in drinking water and OB corresponds to maximum benefits in terms of reduction of ID if asbestos level were reduced to zero.

As benefits are measured in terms of reduction in ID, RID, the graph on the preceding page is converted so as to express the relationship between the reduction in ID and the reduction in asbestos (RAS) level.



As the reduction in asbestos level (RAS) increases, the reduction in ID, RID decreases. Note that beyond some level, reducing the asbestos level does not result in significant reductions in the ID.

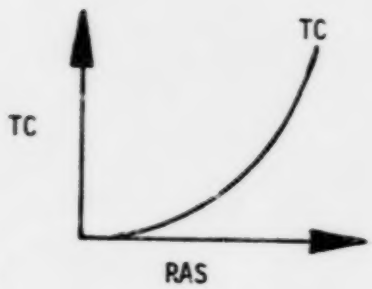
b. Determine TB as the product of ID and V. The pertinent relationship is graphically similar to the above; RID is expressed as the value of total benefits, TB, attained, as reductions in asbestos increase.



Note that TB increases at a decreasing rate as increasing amounts of abestos are eliminated from water. TB becomes horizontal where increasing reductions in asbestos can no longer reduce ID significantly.

The result of a policy changing from standard S_1 to S_2 would provide an increase in total benefits from b_1 to b_2.

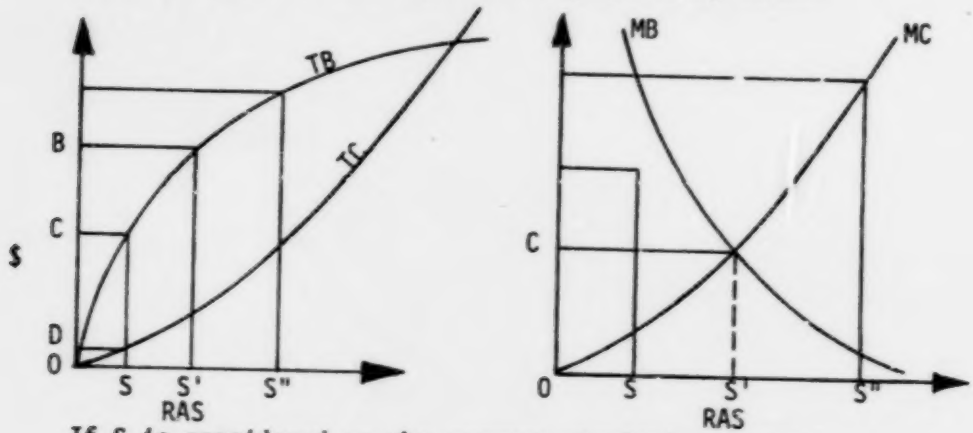
3. Determination of the cost function. The cost function relates a quality standard to minimum cost of achievement. Costs include surveillance, treatment, operators, etc.



Total costs (TC) increase at an increasing rate as the level of reduction in asbestos increases due to the expense of meeting more restrictive standards.

TC becomes vertical if increasing the restrictiveness of a standard (reduction of asbestos level) is no longer possible given the resources (technology and equipment) available.

4. Comparison of benefits and costs. The comparison between TB and TC and between marginal benefits and costs (MB and MC) are represented graphically below.



If S is considered as the present standard, the net benefits to society are represented by DC. The optimum reduction in asbestos level would be attained with standard S', where the vertical distance between TC and TB is greatest. The EPA's "Estimates of Benefits and Costs of the Safe Drinking Water Act of 1973" indicates that water quality standards should be increased. Economic theory tells that the benefits to society from increasing water quality will be maximized where marginal costs equal marginal benefits, MB = MC, at S'.

If marginal (additional) benefits for reducing the level of asbestos through more restrictive standards, exceed marginal (additional) costs, then the restrictiveness of the standard should be increased from S to S'.

On the other hand, if MC < MB, then the restrictiveness of the standard should be decreased from S" to S'.

The benefit/cost comparison requires the determination of both the benefit and cost function described in Step 2 and 3. The process described above refers not only to finding optimum standards for contamination levels, but also to optimum standards for other aspects of water quality control, such as operator training and capital investment.

Procedures

These concepts were used with some modification to develop a benefit/cost analysis for improvements in drinking water programs for a group of three diseases--salmonellosis, shigellosis and hepatitis. Modifications in the process were required because the data for the determination of the incremental cost per unit of change in quality were not available. The following gives an outline of steps as related to sections of this report:

1. Measure the total disease and cost per incidence of disease, C_i (Section IV)
2. Estimate regression equation to explain the reported rate of disease, D_i, and water quality, Q (Section V).
   a. $D_i = f$ (Q, socio-economic variables)
   b. $Q = g$ (source, surveillance, treatment, training)
   c. For a given water quality surveillance treatment and training improvement program, estimate ΔQ, the change in Q, using equation (b). Then use ΔQ in (a) to estimate ΔD_i, the change in reported disease.
3. Costs of a given water quality improvement program will be estimated (Sections VI and VII).
4. From (1) and (2), determine benefits to infectious disease of improvement programs by summing:

   $$\sum_i C_i \, r_i \, \Delta D_i$$

   where r_i is a multiplier (reported to actual disease ratio; Section VII).
5. Determine benefit/cost ratios (Section VII).

The same general analysis technique was applied to an analysis of the effect of several factors, including total hardness, on cardiovascular disease. Different base data were used. This analysis is reported in Appendix D.

SECTION IV
EXTENT OF WATER QUALITY PROBLEM: MEASURING THE COSTS OF INFECTIOUS DISEASE ASSOCIATED WITH DRINKING WATER

Introduction

In order to establish benefits which may be anticipated in response to improvements in the quality of drinking water supplies throughout the United States, estimates must be made of the present costs of disease which can be alleviated by making improvements in the quality of the water delivered to the consumer. The clearest epidemiologic evidence of disease attributable to drinking waters pertains to infectious diseases of protozoan, bacterial and viral origin which includes records of investigations of numerous individual outbreaks as well as the dramatic reduction in the incidence of typhoid fever which followed the widespread chlorination of potable waters. It is the purpose of the present discussion to examine the costs of infectious diseases associated with drinking waters in the United States.

Diseases which may be caused by waterborne infectious agents include the following:

- Salmonellosis (excluding typhoid fever)
- Shigellosis
- *E. coli* Dysentery
- Typhoid Fever
- Leptospirosis
- Tularemia
- Infectious Hepatitis
- Gastroenteritis (non-specific)
- Amebiasis
- Giardiasis

The analysis presented in the following discussion is restricted to three categories of infectious diseases in this list: Salmonellosis, shigellosis, and infectious hepatitis (hepatitis A). These are diseases affecting significant proportions of the total population and data are accessible which can aid greatly in predicting costs. Analyses of costs of others diseases in the list are not included for a variety of reasons. Non-specific gastroenteritis, while it affects many people and is known frequently to be waterborne, is not a reportable disease and ordinarily is of such a mild nature that the extent of the problem is not at all understood, data being almost totally lacking. Other diseases are of minor significance compared with those selected for analysis.

For instance, during 1972, there were reported to Federal Agencies:

Table 1. Reported Infectious Disease

	Total Reported Cases in 1972*	Waterborne Outbreaks 1971-72** Outbreaks	Cases
Salmonellosis (excluding typhoid fever)	22,151	-	-
Shigellosis	20,207	6	692
E. coli Dysentery	1,560	-	-
Typhoid Fever	398	1	5
Leptosporosis	41	-	-
Tularemia	152	-	-
Infectious Hepatitis	54,074	11	266
Gastroenteritis	-	22	5,615
Amebiasis	2,199	-	-
Giardiasis	1,558	3	112

* U.S. Public Health Service, 1973
** Craun and McCabe, 1973b

Although approximately 2,200 cases of amebic dysentery were reported, none were known to be waterborne. There were reported about 1,560 cases each of Giardia infections and E. coli dysentery. The majority of Giardia infections were not of the gastrointestinal tract, and the E. coli infections were largely foodborne. The numbers of reported leptospirosis and tularemia cases were each small and there is no reason to believe that water played a role in their genesis. The number of cases of typhoid fever also was low. Based on data for 1971, one might predict that about 13 percent, or about 58 cases, of the 398 cases reported may have been waterborne. The costs of this small number are minor compared with the costs of salmonellosis, shigellosis and infectious hepatitis.

Although no waterborne salmonellosis was reported in 1972, data for previous years provide a basis for analysis, as do the data for shigellosis, and while the analysis of costs associated with salmonellosis shigellosis are crude, they are better than would be similar analyses of costs associated with other infectious diseases. Whereas the costs associated with waterborne infectious hepatitis are perhaps less well founded, the disease is one of substnatial importance and therefore is included in the analysis.

Estimates of the incidence of infectious disease associated with drinking waters are based upon reported cases with limited studies of the reporting of disease. Costs connected with such disease are based upon estimates of time lost from work and time spent in hospitals. No attempt is made to assess costs of investigations of outbreaks. Nor are estimates made of secondary cases beyond those which are included in the raw data. Secondary cases can represent a significant number

of illnesses. However Lobel, et al. (1969), investigated secondary spread of a waterborne non-specific gastroenteritis among 64 family contacts of 65 patients in 61 families. Twenty-eight family contacts in 14 households became ill. On the other hand, although infectious hepatitis is felt to be spread commonly through person-to-person contact, Garibaldi, et al. (1972), reported secondary cases in only 6 household contacts of 79 patients who had become ill as a result of drinking contaminated water in a restaurant. No efforts were made to account for more elusive costs related to suffering associated with either illness or death.

Comparisons are made with data on costs of the same diseases obtained from other sources. These include an analysis of the Salmonella problem by the National Academy of Sciences (Committee on Salmonella, 1969), a review of turtle associated salmonellosis (Lamm, et al., 1972), an estimation of costs associated with viral hepatitis (Tolsma and Bryan, unpublished), and EPA estimates of costs associated with waterborne disease (Enviornmental Protection Agency, 1973).

Estimates of Disease Incidence

Salmonellosis. The number of Salmonella outbreaks (exclusive of typhoid outbreaks) reported during any year is highly variable. Thus an average yearly number of reported outbreaks and associated cases of waterborne Salmonella infections over the ten-year period, 1961-70, reported by Craun and McCabe (1973a) has been employed as a basis for estimating the actual number of yearly cases.

Over the ten-year period, 1961-70, there were reported by Craun and McCabe, 9 outbreaks of salmonellosis involving 16,706 cases. Of the outbreaks, only one-third had been reported to the Center for Disease Control. The reliability of reporting of outbreaks can be judged further on the basis of intensive surveillance of foodborne epidemics in the State of Washington (Barker, et al., 1969). During 1969 the incidence of foodborne outbreaks in the State of Washington was 23 per million population, as opposed to a rate of 1.9 outbreaks per million population reported to the Center for Disease Control. These figures suggest a ratio of actual to reported outbreaks amounting to at least 12:1. Furthermore, the relationship between the reporting of salmonellosis and other foodborne diseases in Washington and the United States was the same as the reporting of total foodborne outbreaks. There is no reason to suspect that the reporting of waterborne disease should differ from the reporting of foodborne disease. Therefore, the actual incidence of water borne salmonellosis is taken to be four times that reported by Craun and McCabe (1973a). (Note: The ratio of 12.1:1 observed by Barker, et al. [1969], divided by the factor of 3:1 observed by Craun and McCabe [1973a].) Thus the data of Craun and McCabe can be adjusted to yield an estimate of 36 outbreaks and 66,824

cases of waterborne salmonellosis over the ten-year period.

Further investigations of the reporting and true incidence of salmonellosis have been undertaken by Aserkoff, et al. (1970). These authors examined the findings of three epidemiologic investigations. The number of reported cases in the three outbreaks was less than one percent of the cases subsequently discovered during surveys of the exposed population. Aserkoff, et al., suggested that, on this basis, the reported 20,000 or so Salmonella infections reported to the Center for Disease Control represented a true figure of about 2,000,000 cases per year in the United States. An estimate can be made of the true number of cases of waterborne salmonellosis by multiplying by 100 the number of cases reported by Craun and McCabe and corrected for reporting of outbreaks. Because 16,000 cases occurring in Riverside represent all cases in that epidemic, this number of cases cannot be multiplied by 100. Thus the true ten-year incidence of waterborne salmonellosis is probably closer to:

$$(16,706 - 16,000) \times 4 \times 100 + 4 \times 16,000 = 346,000$$

or 34,640 cases per year. If the total yearly cases of salmonellosis is 2,000,000, waterborne disease appears to comprise only about 1.7 percent of the total cases.

Shigellosis. The extent of waterborne shigellosis may be estimated in a manner similar to that of waterborne salmonellosis. The number of reported outbreaks of shigellosis was greater than that of salmonellosis (Craun and McCabe, 1973a): 19 outbreaks and 1,666 cases. Since outbreaks reported by Craun and McCabe nearly always had been reported to the Center for Disease Control, the ratio of actual to reported cases suggested by the data of Barker, et al. (1969), may be applied directly to estimate the actual number of outbreaks: 230 outbreaks affecting 20,150 people over the ten-year period. If it is assumed that the reporting of cases of shigellosis, like that of salmonellosis is about one percent of the total, the actual incidence of waterborne shigellosis might be about 201,500 cases per year.

The average number of Shigella isolations from human sources reported to the Center for Disease Control is about 9,000 per year (Reller, et al., 1970), but to our knowledge no estimates of total cases have been made. In contrast to salmonellae, shigellae are not associated strongly with animals but rather are associated primarily with man. Thus, proportionately more shigellosis than salmonellosis may be related to drinking water, and the estimate of 201,500 cases per year is perhaps not unreasonable.

Infectious hepatitis. There were reported during the year 1972, 54,074 cases of acute infectious hepatitis and hepatitis of unspecified type (Center for Disease Control 1973a). During the epidemiologic year, 1972, histories of 15,447 cases of infectious hepatitis indicated that

1.1 percent had received a transfusion, 10.8 percent had a history of parental drug abuse, and 30.4 percent had had personal contact with known cases. In the investigations of the Holy Cross outbreak (Morse, et al., 1972a, b; Wacker, et al., 1972), it was found that 56 percent of those exposed exhibited symptoms. An additional 37 percent exhibited abnormal liver function (elevated serum glutamic pyruvic transaminase, SGPT) but were free of symptoms. Applying these figures, the estimate of known personal contact has been increased by a factor of 1.66 ([56% + 37%] ÷ 56%) to account for exposure to asymptomatic carriers. Thus 50.5% (1.66 x 30.4%) can be assumed to have had personal contact with either symptomatic or asymptomatic cases. Consequently, no potential source of infection was evident in the histories of 37.6 percent (100% - 50.5% - 10.8% - 1.1%) of the patients having infectious hepatitis.

The reported cases of acute infectious hepatitis due to water can be estimated by correcting for cases of unknown etiology. Among cases of infectious hepatitis and those of unspecified type for which histories were examined at the Center for Disease Control, 88.7 percent hepatitis A. Thus, about 48,000 cases (88.7% of 54,074 cases) may be considered to be infectious hepatitis, of which 18,000 (37.6%) were of unknown genesis.

It has been reported (Morse, et al., 1972a, b; Wacker, et al., 1972) for the Holy Cross hepatitis outbreak that of persons exposed to the contaminated water, 56 percent were symptomatic. Of those ill, 59 percent were icteric and 41 percent anicteric. Applying these figures to the cases reported with no known exposure, it can be estimated that about 10,600 cases of icteric infectious hepatitis and 7,400 cases of anicteric symptomatic infectious hepatitis occurred.

It can be assumed safely that essentially all icteric (jaundiced) cases reported their disease to a physician. Based upon interviews with physicians, however, it appears safe to assume that only about 50 percent of anicteric (non-jaundiced) hepatitis patients reported to a physician. Thus, it appears reasonable to adjust the number of anicteric symptomatic cases by multiplying by a factor of 2, yielding an adjusted estimation of the incidence of infectious hepatitis (29,400 total cases, 10,600 icteric cases, and 14,800 anicteric symptomatic cases). It is known, furthermore, that not all known cases of hepatitis are reported by the physician. Even where hospital charts are reviewed conscientiously for that purpose, only 15 to 40 percent are reported to the Professional Activity Study (PAS) of the Commission on Professional and Hospital Activities (CPHA) (Eichoff, et al., 1969). Thus, if 25 percent of the total hepatitis cases are assumed to be reported and included in the national figures, the true extent of infectious hepatitis cases having no known potential contact with the disease must be considered to be the following:

Icteric	42,400
Anicteric Symptomatic:	
Seeing Physician	29,600
Not Seeing Physician	29,600
Total	101,600

It is assumed further that the cases for which no potential genesis is known were food and waterborne cases. To estimate how many of these cases might be waterborne, a relationship between food and waterborne cases of infectious hepatitis must be employed. Such a relationship may be derived from data reported in the Annual Summary of Foodborne 1972 (Center for Disease Control, 1973b). During the 1971-72 year there were reported to the Center for Disease Control 11 outbreaks of waterborne infectious hepatitis including 266 cases. There were reported for the same year 5 foodborne outbreaks including 90 cases. Thus, the ratio of waterborne to foodborne cases is considered to be waterborne. Applying this figure to the previous approximation of foodborne and waterborne cases, it is estimated that the yearly occurrence of waterborne infectious hepatitis is as follows:

Icteric	31,800
Anicteric Symptomatic:	
Seeing Physician	22,200
Not Seeing Physcian	22,200
Total	76,200

The estimated number of waterborne infectious hepatitis cases is again very much higher than the reported cases. This is not out of line with the estimate of a total of 570,000 cases in 1970 made by Tolsma and Bryan (unpublished). If the above percentages are applied to the estimate of total cases made by Tolsma and Bryan, the yearly total number of waterborne infectious hepatitis cases might amount to 142,700 (37.6% x 88.7% x 75% x 570,000), or nearly twice the cases estimated by applying the present procedures. Over the period 1961-70 Craun and McCabe (1973a) were able to identify 30 waterborne outbreaks involving 903 cases, only about one-half of which had been reported to the Center for Disease Control. Thus, the yearly number of reported cases was about 90 between 1961 and 1970, and the cases reported in 1971 were 266, or 0.12% and 0.35% of the estimated cases, respectively.

Costs Associated with Lost Time

In order to justify categorical program expenditures, costs often are estimated in gross terms, with emphasis on the economic impact of the disease in question on society. In gross estimations of costs of illness, average earnings commonly are applied to overall morbidity data,

often with no adjustments for labor force participation and unemployment rates. Furthermore, age and sex differences in earnings may or may not be taken into account, public payments and taxes on earnings may not be accounted for properly, and direct costs may consist of a conglomeration of educated guesses and available statistics on institutional care, negelecting noninstitutional costs (Rich, 1966). The present analysis of costs of waterborne illness is of this type.

Days lost. Little note is made in the literature of length of illness, which is perhaps a little surprising in view of the attention often devoted to the frequencies of symptoms. On the other hand, contact with the physician ceases when the patient no longer requires his services, and the length of illness is not necessarily known to the physician. However, time lost from work and length of illness would appear to be significant points in follow-up surveys in epidemiologic studies. Some data are available, and it is possible to estimate lengths of illness for the infectious diseases under consideration.

Both salmonellosis and shigellosis are diseases the acute phases of which are of relatively short duration. Bauer (1973) has stated that recovery from Salmonella infections occurs in 2 to 5 days. Although Cherubin, et al. (1969), state that the mean duration of symptoms is 11.8 days, the former is consistent with estimates of lost time given in interviews with practicing physicians, and 4 days is taken as the mean length of time lost from productive activities. Symptoms of shigellosis linger, sometimes for long periods of time. For instance, investigations of an outbreak of Shigella flexneri 2a, among members of a tour to Mexico indicated that the duration of illness was less than 5 days in 34 percent of the cases, 6 to 10 days in 29 percent, 11 to 15 days in 19 percent, and greater than 15 days in about 17 percent (Center for Disease Control, 1971). However, the disease is not as disabling as is salmonellosis. The length of time lost as a result of shigellosis also is taken as 4 days. Furthermore, this is the average length of time spent in the hospital by hospitalized patients (Barrett-Connor and Connor, 1968).

The problem, of estimating length of time lost as a result of infectious hepatitis, is less easily settled. Interviews with practicing physicians indicate that patients visit their physician after about 7 days of illness. All patients contacting a physician enter the hospital for tests which require hospitalization for about 7 days (the geometric mean length of stay estimated from figures supplied by the Professional Activity Study of the Commission on Professional and Hospital Activities is about 7.5 days). Patients with jaundice may be expected to lose about 21 days. Comparison with published information on length of illness supports the present estimates. Data published by Krugman, et al (1962), indicates that tests of liver function return to normal about 28 days after onset of illness. Repsher and Freebern (1969) published data suggesting that the length of illness was slightly in excess of 4 weeks, although in a summary of their

own data and those of others studying military populations, they indicated that jaundice may last from 14 to 35 days, and that durations of hospital stays were in excess of 5 days. The latter figure is consistent with data of Andrews (1970) which indicated that the average time away from work was 7 1/2 weeks. Thus, the assumed lengths of illness are not excessive. One final assumption is that the population of infectious hepatitis cases not seeing a physician is composed of mild cases, and that these will lose, on the average, 7 days of normal activity.

To aid in assessing economic value of time lost, a mean daily income has been employed. To determine a mean daily income, we have used the mean per capita income for the United States before taxes, but after deductions for social security, government retirement and other social insurance programs (U.S. Department of Commerce, 1973). This amount, which in 1971 was $4,157, has been divided by 365 to yield a daily per capita income of $11.40.

Costs. Employing the above estimates of lengths of illness and per capita income, along with previous estimates of the incidence of waterborne disease, costs associated with loss of time have been predicted:

Table 2. Annual Costs of Illness (Lost Time)

Disease	Incidence*	Length of Illness	Days Lost	Total Cost of Illness
Salmonellosis	35,000	4	140,000	$ 1,600,000
Shigellosis	202,000	4	808,000	9,200,000
Infectious Hepatitis				
Icteric	32,000	30	960,000	10,950,000
Anicteric Symptomatic				
Seeing Physician	22,000	21	462,000	5,250,000
Not Seeing Physician	22,000	7	154,000	1,750,000
Total				$28,750,000

* Rounded off to nearest 1,000

Costs of Hospitalization

Of all medical expenses associated with waterborne infectious disease, only those costs associated with hospital stays are considered significant and thus are included in the present analysis of costs.

Hospitalization of patients with Salmonella and Shigella infections is rare. Projections based upon data collected in the PAS system indicate

that between 5,000 and 6,000 primary cases of each disease were hospitalized during 1972. If the reporting of cases 40 percent on the average (Eichoff, et al., 1969), then the actual number of total hospitalized cases of salmonellosis and shigellosis each might be 12,000. The number of hospitalized cases of salmonellosis related to water can be estimated in several ways. First, if it is assumed that the proportion of total cases hospitalized, then

$$\frac{\text{Total Number Hospitalized}}{\text{Total Number Ill}} = \frac{\text{Waterborne Cases Hospitalized}}{\text{Waterborne Cases}}$$

or

$$\frac{12{,}000}{2{,}000{,}000} = \frac{\text{Waterborne Cases Hospitalized}}{35{,}000}$$

and 210 cases of waterborne salmonellosis might have been hospitalized. Alternatively, figures of Aserkoff, et al. (1970), for numbers ill and hospitalized in 180 epidemics adjusted for reporting can be employed. In this way:

$$\frac{\text{Number Reported Hospitalized Corrected for Reporting (40\%)}}{\text{Number Ill Corrected for Reporting}} = \frac{\text{Waterborne Cases Hospitalized}}{35{,}000}$$

$$\frac{645 \times 2.5}{16{,}772 \times 100} = \frac{\text{Waterborne Cases Hospitalized}}{35{,}000}$$

and only 605 cases of waterborne salmonellosis might have been hospitalized. It is probably reasonable to choose an intermediate figure of 400.

While the number of waterborne shigellosis cases probably was higher than the number of salmonellosis cases, there is no basis for estimating such numbers and the figure is taken to be 400 also.

We have indicated above that the average hospital stay of patients with shigellosis has been reported to be 4 days (Barrett-Connor and Connor, 1968). There are no data available on lengths of stay from the Professional Activities Study since such analyses are not made on samples of less than 30,000 cases. Thus, we make the assumption that hospital stays for both diseases are 4 days. This appears to be consistent with data obtained during epidemics of salmonellosis. In an analysis of outbreaks of turtle-associated salmonellosis, Lamm, et al. (1972), reported average hospital stays of 2.8 days. In an outbreak of S. infantis gastroenteritis (Committee on Salmonella, 1969) an average hospital stay of 5.5 days was reported.

The length of hospital stays of patients with infectious hepatitis is more easily determined. Clearly, those patients not contacting their physician have no hospital expenses. A joint study of the Commission on Professional and Hospital Activities and LEA, Inc. Commission on Professional and Hospital Activities (1973) provides an analysis of hospitalized cases of infectious hepatitis. Based upon the distribution of lengths of stays, the geometric mean stay in 1972 was 7.5 days and the arithmetic mean stay was 10.8 days. The former figure is applied to both icteric and anicteric symptomatic cases.

Costs associated with hospitalization are elusive. Probably about the best estimate that can be made is that based upon average semi-private hospital room and board charges published by the Equitable Life Assurance Society of the United States. In 1973 the average cost for the country as a whole was $59.50 per day. As a rule, charges for physician and other services approximate room and board charges (Mr. J. B. Ramage, Equitable Life Assurance Society of the United States, Personal Communication). Thus the daily costs during hospitalization are taken to be $120. It is worth noting also that costs of hospitals and hospital care increase at about 14 percent per year (the Conference Board, No. 1721, September 1, 1973).

Based upon the above estimates of numbers and lengths of stays in hospitals and costs of hospitalization, the following costs associated with waterborne infectious disease have been predicted:

Table 3. Annual Costs of Hopitalization

Disease	Hospital Cases	Length of Stay	Total Days of Hospitalization	Total Cost of Hospitalization
Salmonellosis	400	4	1,600	192,000
Shigellosis	400	4	1,600	192,000
Infectious Hepatitis				
Icteric	32,000	7.5	240,000	28,800,000
Anicteric Symptomatic				
Seeing Physician	22,000	7.5	165,000	19,800,000
Not Seeing Physician	22,000	-	-	-
Total	76,800		408,200	48,984,000

Costs Associated with Loss of Life

Although death rates for waterborne infectious diseases are low, they may constitute a significant part of the costs associated with the diseases. In 1972, the overall fatality ratio for infectious hepatitis was 0.7 percent, and that observed for salmonellosis outbreaks 0.93 percent (U.S. Public Health Service, 1973; Center for Disease

Control, 1973a). While mortality figures are not available from the Shigella Surveillance reports of the U.S. Public Health Service, Barrett-Connor and Connor (1968) reported a mortality rate of 1.5 percent among 330 confirmed cases at Jackson Memorial Hospital. These rates cannot, however, be applied to the total estimated waterborne disease. Rather, they can be applied safely only to reported cases. Thus, the mortality rate of 0.93 percent must be applied to 16,706 reported cases of waterborne salmonellosis over 10-year period, or 1,670 cases and 16 deaths per year. Similarly, the mortality rate of 1.5 percent must be applied to 20,150 reported cases of waterborne shigellosis over a 10-year period, or 2,015 cases and 30 deaths per year. Deaths of infectious hepatitis cases occurred among 0.75 percent of the 18,000 cases of unknown genesis. Thus, approximately 95 deaths per year can be attributed to waterborne infectious hepatitis.

The value of an average life calculated from estimated deaths and costs of mortality in 1970 presented by Tolsma and Bryan (unpublished) was $64,000, and represents lost earnings. The following cost associated with mortality caused by waterborne infectious diseases have been estimated:

Table 4. Annual Costs of Mortality

Disease	Reported Cases	Mortality Rate (%)	Number of Deaths/Year	Total Cost of Mortality
Salmonellosis	1,670	0.93	16	1,024,000
Shigellosis	2,015	1.5	30	1,920,000
Infectious Hepatitis	13,500	0.7	95	6,080,000
Total	17,185		141	9,024,000

Costs Associated with Production Losses

In order to estimate costs associated with production losses and losses of generated income, a per capita gross national product was determined for the year 1967:

$$\text{GNP/Labor Force} = \frac{793.9 \times 10^9}{80.1 \times 10^5} = \$10,000$$

An output multiplier of 1.8 was applied to the per capita GNP to give a per capita generated income of $18,000. The multiplier was chosen as typical of most occupations and a reasonable average value.

Assuming 250 workdays per year, the daily loss of generated income was estimated at:

$$\frac{\$18,000}{250} = \$72/\text{day}$$

An estimate of the number of workdays lost was made by adjusting the days lost used to determine costs associated with lost time. The number of lost days was adjusted by a factor accounting for the length of the work week and was applied only to that portion of the population of the United States in the labor force between 20 and 64 years of age (52.25%):

$$\text{Factor} = \frac{.5225 \times 5}{7} = 0.373$$

Applying the preceding equation, daily loss of generated income and the adjustment factor, the following costs have been assigned to the waterborne diseases under analysis:

Table 5. Annual Costs of Generated Income Lost

Disease	Work Days Lost	Loss of Generated Income
Salmonellosis	52,300	$ 3,760,000
Shigellosis	328,000	23,600,000
Infectious Hepatitis		
Icteric	358,000	25,800,000
Anicteric		
Seeing Physician	172,000	12,390,000
Not Seeing Physician	57,500	4,140,000
Total		$69,690,000

Total Costs of Infectious Disease Associated with Drinking Waters

Total costs of waterborne infectious diseases are summarized in the final table. The total cost is estimated to be $156,448,000, of which $28,750,000 is associated with lost work time, $48,984,000 is attributable to hospital costs, $9,024,000 to mortality losses, and $69,690,000 to production losses. The largest costs are those associated with infectious hepatitis ($17,950,000 for lost time and $48,600,000 for hospitalization) with a total cost of $114,960,000. Thus, the accuracy of the estimates of costs of waterborne diseases depend heavily upon the accuracy of the estimates of numbers of infectious hepatitis cases associated with drinking waters.

Costs of Waterborne Disease According to Type of System and Cause of Outbreaks

An estimate of the costs of waterborne disease associated with public and private systems and cause of outbreak have been estimated using tables presented by Craun and McCabe (1971) for the years 1961-1970. According to this analysis, the major costs are associated with public

untreated ground water supplies and inadequate control of treatment in public supplies ($65,100,000 and $62,800,000, respectively, out of a total of $156,448,000).

Table 6. Distribution of Costs of Three Infectious Diseases by Cause and Type of System (Annual Costs)

Cause	Private Systems	Public Systems	All Systems
Untreated Surface Waters	666,000	1,202,000	1,868,000
Untreated Ground Waters	6,865,000	65,235,000	72,100,000
Contamination of Storage Facilities	437,000	-	737,000
Inadequate Control of Treatment	5,591,000	62,914,000	68,505,000
Contamination of Distribution System	219,000	4,406,000	4,625,000
Miscellaneous	8,112,000	801,000	8,913,000
Total	21,890,000	134,558,000	156,448,000

However, the assignment of costs cannot be made properly, in this way, since Craun and McCabe's distribution of cases includes gastroenteritis and other diseases not included in the present analysis. Furthermore, it is based upon reported cases only, which may differ substantially from the distribution of total cases estimated in the present analysis.

A similar analysis of the distributution of cases of infectious hepatitis based upon reported cases between 1946 and 1970 (Craun and McCabe, 1973a) has associated with it the question of distribution of actual and reported cases, and in addition includes cases reported in the 1940's and 1950's, which also may have been distributed differently from more recent cases. However, it has the advantage of employing data for infectious hepatitis only. No similar data for salmonellosis and shigellosis have been compiled.

The greatest costs associated with waterborne infectious hepatitis may be attributed to public systems using untreated ground waters ($46,150,000 out of a total of $114,900,000). Furthermore, while approximately 40 percent of costs associated with waterborne infectious hepatitis occurs among persons served by private systems, only about 14 percent of the total costs associated with salmonellosis, shigellosis and infectious hepatitis occurs among persons served by private systems.

Table 7. Distribution of Costs of Infectious Hepatitis by Cause and Type of System (Annual Costs)

Cause	Private Systems	Public Systems	All Systems
Untreated Surface Waters	11,850,000	8,440,000	20,290,000
Untreated Ground Waters	970,000	46,150,000	47,120,000
Inadequate or Interrupted Disinfection	5,360,000	9,550,000	14,910,000
Contamination of Distribution System	18,580,000	3,020,000	21,600,000
Unknown	9,540,000	1,440,000	10,980,000
Total	46,300,000	68,600,000	114,900,000

Comparison of Costs with other Analyses

Estimates have been reported of total costs associated with salmonellosis in the United States and of the costs of turtle-associated salmonellosis in the United States. Although the bases for making estimates differ from those in the present study, it is useful to make the comparison. The Committee on Salmonella of the National Academy of Sciences (1969) estimated costs to patients of 2,000,000 annual cases at between $20 million and $200 million. Applied to waterborne cases, this would be equivalent to $350,000 to $3,500,000 per year. The cost of $1,648,000 estimated in the present study lies safely between these figures. The more recent analysis of turtle-associated salmonellosis by Lamm, et al. (1972) sets the average cost of medical expenses at more than $300 per case. This cost is substantially higher than that employed in the NAS analysis and in the present analysis. But the population of turtle-associated cases was a young one and not typical of waterborne salmonellosis, or the national picture as a whole.

Tolsma and Bryan (unpublished manuscript) have approximated the costs of an estimated 570,000 annual cases of acute infectious and non-specific hepatitis employing more rigorous procedures than were applied in the present analysis. Their estimated costs included $115.5 million in direct costs and $227.7 million in costs attributable to death and lost time. This amounts, on the average, to a cost of $600 per case. Applied to the estimate of 76,000 total waterborne cases in the present analysis, this would lead to an estimate of $45,600,000, which is about 29 percent of the cost ($114,940,000) estimated in the present study. On the other hand, the estimate of numbers of cases in this study was far more conservative than that of Tolsma and Bryan, who also did not take into account production losses. Thus the estimate is not considered out of line.

Table 8. Annual Costs of Infectious Disease Associate with Drinking Water

Disease	Lost Work Time	Hospitali-zation	Mortality	Production Losses	Total Costs
Salmonellosis	1,600,000	192,000	1,024,000	3,760,000	6,576,000
Shigellosis	9,200,000	192,000	1,920,000	23,600,000	34,913,000
Infectious Hepatitis			6,080,000		6,060,000
Icteric	10,950,000	28,800,000	-	25,800,000	65,550,000
Anicteric Symptomatic					
Seeing Physician	5,250,000	19,800,000	-	12,390,000	37,440,000
Not Seeing Physician	1,750,000	-	-	4,140,000	5,890,000
Total	28,750,000	48,984,000	9,024,000	69,690,000	156,448,000

The figures obtained in the present study differ substantially from those compiled by EPA in an earlier unpublished report (Safe Drinking Water Act of 1973; Estimates of Benefits and Costs). The three diseases considered here are included, along with typhoid fever and amebiasis, under category two of that report. The total costs of the 5 diseases were placed at $1.2 billion per year. While we consider the estimates made in the present study to be on the conservative side, the true costs cannot be considered to approach the EPA figure, and the EPA figure appears inflated. Indeed, the EPA figure may exceed the total yearly costs of the five diseases in question in the United States. For instance the NAS estimates of the costs of total salmonellosis are between $20 million and $200 million annually, and Tolsma and Bryan judged the costs of all infectious hepatitis to be $343.2 million annually.

SECTION V
METHODOLOGY FOR ANALYSIS OF BENEFITS OF DRINKING WATER IMPROVEMENT PROGRAMS

A. Purpose of Analysis and Techniques of Measurement

The research effort for this project includes two types of benefit measurements related to drinking water:

1. Measurement of the extent of losses due to waterborne disease (total social cost)
2. Measurement of benefits of specific water quality improvement programs (reduction in social cost)

The previous chapter addressed itself to the first question; this chapter will relate mainly to the second. While the benefits of a specific program cannot exceed the value of total losses due to waterborne diseases, any particular improvement program cannot be expected to eliminate all waterborne disease. The amount of waterborne disease averted by a program obviously depends on the nature and extent of the program. Measurement of benefits of specific water quality improvement programs thus requires a method of analysis based on specific program parameters and how they are related to disease. This chapter will present such a methodology.

As an offshoot of this methodology it is also possible to estimate the number of cases of reported disease actually due to water and hence to measure the losses due to waterborne disease in a different way than that in Section IV.

In Section IV the number of cases of disease due to water were estimated by taking the cases which were reported to the CDC as being definitely caused by water and using a multiplier to account for unreported disease; this method may exclude secondary cases. In this section, we start from <u>total</u> reported disease and seek to identify which part of the total may be traced to water either directly or indirectly. A multiplier (different from the ones used in Section IV) may then be used to adjust for non-reporting.

The method will be applied to measure benefits associated with reduction of three water related diseases (hepatitis, shigellosis and salmonellosis) due to a specific improvement program. The specific program to be analyzed consists of the following parameters:

1. That all community water supply systems meet current EPA surveillance criteria
2. That all systems maintain a total residual chlorine of 0.3-0.5 ppm
3. That in each system there is adequate disinfection capacity

4. That each system serving at least 1,000 to 500,000 people have a full time operator and all others have at least a part-time operator

The methodology could be used to study other diseases and other programs. However, the choice of diseases was limited here by the amount of information available from public health services and the Center for Disease Control (CDC) and the choice of a program was limited by the amount of information available from the Community Water Supply Study (CWSS). The above program, however, is a reasonable one and is consistent with EPA recommendations.

The method of analysis will be based on the development of two types of relations. The first type, termed disease equation, will relate reported rates of disease to water quality and socio-economic variables. A separate equation will be developed for each of the three diseases since the relation of disease to water quality and socio-economic variables can be expected to vary depending on the disease. The second type of relation will be termed a water quality production equation and will relate water quality to water system characteristics (treatment process characteristics, raw water quality, operator characteristics and surveillance). The two types of relations will be used in combination to measure reduction in disease due to a program.

To develop the relations, a technique called multiple regression analysis is used; an explanation of this technique can be found in many texts on econometrics (Goldberger, 1964; Kmenta, 1971). The technique essentially is a statistical means of estimating parameters relating independent variables to a dependent variable to obtain the equation with the best "fit" to the data. A separate regression equation is estimated for each of the three diseases studied here.

Regressions are made using three sets of data: data from the CWSS on water quality and treatment, data from the U.S. Census on socio-economic variables, and disease data obtained directly from individual counties. For the disease equations cross sectional data used in the regression was for 39 out of 42 counties included in CWSS; county boundaries were used because these most closely matched the boundaries of the three data sets. For the water quality production equation, data used was for only 870 of the 969 in the CWSS due to inconsistencies in the data. An appendix gives more detailed data descriptions and discussion of data problems.

As discussed by Kmenta (1971), besides estimating a relationship, regression can be used for several purposes:

1. To test hypotheses (eg. whether or not water quality affects rate of disease, whether or not residual chlorine affects water quality in the distribution system, etc.). (Note: A hypothesis about a regression coefficient is tested by means of a t-statistic which relates to the probability of a regres-

sion coefficient being different from zero. For example, if the t-value for a coefficient is greater than 1.28, then the hypothesis that the coefficient is zero is rejected at the 90 percent level.)

2. For prediction (what will be the water quality for a given treatment program, what will be the rate of disease for a particular water quality and set of socio-economic variables, etc). (Note: A prediction is made by using a set of independent variables in the regression equation and obtaining the corresponding value of the dependent variable. A standard error (s.e.) is used to obtain a confidence interval depends on a probability parameter and the degrees of freedom,

$$Pr[Y_p - t_{\alpha/2}\ (s.e.) \leq Y \leq Y_p + t_{\alpha/2}\ (s.e.)] = \alpha$$

where Y_p denotes the predicted value and Y denotes the true value. The values of $t_{\alpha/2}$ are found in a table of t-statistics depending on the degree of freedom.)

3. To measure the relative importance of variables in the regression (how important is water quality compared to socio-economic variables in explaining disease; what is the relative importance of the various program parameters). (Note: Relative importance is compared using a "beta coefficient" as defined in Goldberg (1964). This involves "standardizing" the variables so that the effect of a unit change in each may be compared. The beta coefficient is:

$$\beta_j = a_j\ (S_{jj}/S_{yy})$$

where a_j is a regression coefficient, S_{jj} is the standard deviation of the corresponding independent variable and S_{yy} is the standard deviation of the independent variable.)

In addition to estimating regression equations, the analysis here will also include the above considerations. The following sections present the results of the analysis.

B. Disease Relations

As mentioned above in the regression equations, the rate of reported incidences of a water-related disease is hypothesized to be related to water quality and socio-economic variables. In the regressions estimated here, the disease variable will be measured in terms of a rate per thousand population by county. This rate figure includes all cases of reported disease (shigellosis, hepatitis or salmonellosis), whether or not the cases have been shown to be caused by water. (Note: The term "water-related" means disease which could be caused by water either directly or indirectly, while "water caused" means disease which is actually confirmed as caused by water.)

It was felt that regression equations based on explaining total reported disease (including water-related disease) would be more accurate to predict water related disease than regressions based on confirmed water-caused disease since there are more cases of disease related to water than are identified as water-caused. While all of the three diseases studied here must be reported to CDC, the cause of the disease may not be determined unless it is considered to be an outbreak. Thus, part of the problem is to estimate how much of all reported disease is actually related to water quality and how much is due to other factors as reflected in the socio-economic variables.

Water quality for the distribution system is used as an independent explanatory variable in the disease regression and is measured as the average Standard Plate Count (number of bacteria/ml) in a county. This measure of water quality, rather than a measure at the treatment plant, is used because it is the quality of water reaching the consumer which influences disease.

Since water quality is not the only source of the diseases studied here, other variables which may explain differences in reported rate of disease among counties must be considered. Actual disease rates may differ among counties due to socio-economic factors and availability of medical services. There may also be differences in disease reporting among counties due to methods of reporting disease.

Several classes of socio-economic variables may be hypothesized to effect disease rates apart from water quality. Since crowded conditions facilitate spread of disease, a measure of population concentration may be used to reflect this aspect. Alternative explanatory variables include the density of population (PPM) percent urban population (URB) and percent population served treated water (PPS). General health levels in an area may also effect the baseline level of disease apart from disease caused by water. An income per capita variable (PPY) may be used to reflect this aspect with the hypothesis that the higher the income level, the better the general health level with respect to infectious diseases. Some age groups are also more susceptible to certain diseases than others. For example, most reported cases of hepatitis occur between the ages of five and twenty-five (CDC, 1972). Thus an age level variable can also be used to explain differences in reported disease among counties. These three types of socio-economic variables, age, income and concentration of population, are used in the regression models to reflect differences in baseline levels of disease among counties. Differences in disease reporting among counties may be due to differences in the emphasis put by a state or county on accurate reporting. Emphasis put by state and counties on disease reporting can be hypothesized to be related to the level of state and county health expenditures per capita (SLPC); the higher the expenditure level is, the better the reporting of disease.

One further set of explanatory variables was used. Measurements of water quality for the CWSS were taken at different times of the year

for different geographic areas. To correct for possible errors of water quality measurement in the laboratory and/or seasonal differences, three dummy variables (T_1, T_2, T_3) are used. The constant term in the regressions are suppressed because of the use of these dummy variables.

Tables 9, 10 and 11 following present the results of the regression equation for each of the three diseases. The R^2 values indicate how much of the variance in reported disease rates are due to the variables in the regression. While quite high R^2 values (0.9 and above) were obtained for regressions explaining numbers of disease cases due to the heavy influence of population variable, R^2 values for the rate equations are not so high due to elimination of population as a variable. This is a common phenomenon when estimating rate versus total equations; the R^2 values obtained are considered quite good for rate equations.

The regressions for rate of hepatitis and shigellosis are linear of the form

$$\text{Rate} = \Sigma_i \alpha_i X_i \tag{1}$$

where α_i is the estimated coefficient of the independent variable X_i. A coefficient α_j has the interpretation of the change in the rate due to a unit change in variable X_j, all other independent variables held constant; thus for a given change ΔX_j with all other variables constant

$$\Delta \text{ Rate} = \alpha_j \Delta X_j \tag{2}$$

The regression for salmonellosis is of the form

$$\text{Rate} = \Sigma_i \alpha_i \log X_i \tag{3}$$

or equivalently

$$e^{\text{Rate}} = \Pi_i X_i^{\alpha_i} \tag{4}$$

The coefficient αj in this case is the change in the rate due to a percent change in X_j, all other variables constant. Thus for a given X_j and change ΔX_j

$$\Delta \text{ Rate} = \alpha j \frac{\Delta X_j}{X_j} \tag{5}$$

(i) Hypothesis testing and importance of variables in the regressions

The same variables are used in the rate equation for hepatitis and shigellosis. As hypothesized, the water quality variable (SPCD), the density variable (PPM) and income variable (PPY) have positive coefficients for both diseases indicating that rate of reported disease

Table 9. Infectious Hepatitis
(Dependent Variable - Rate of Infectious Hepatitis/1,000 Population)

Independent Variable	Estimated Coefficient	T-Value	β-Coefficient
SPCD	.0001601	1.449	.337
PPM	.000007059	.846	.1794
SLPC	.0107	.4584	.1239
PFTF	2.299	1.8456	.3856
PPY	.00007803	1.4903	.3011
PPS	-.1207	-.7621	-.1915
T_1	-.9223	-1.599	-
T_2	-.9672	-1.7016	-
T_3	-.8660	-1.5121	-

R^2 = 0.3021
Degrees of Freedom = 30

Table 10. Shigellosis
(Dependent Variable - Rate of Shigellosis/1,000 Population)

Independent Variable	Estimated Coefficient	T-Value	β-Coefficient
SPCD	.0001104	2.612	.535
PPM	.000003784	1.185	.218
SLPC	-.001508	-.1674	-.040
PFTF	.9361	1.9636	.361
PPY	.00002276	1.1362	.202
PPS	-.007499	-.1236	-.027
T_1	-.4455	-2.0186	-
T_2	-.4310	-1.984	-
T_3	-.3754	-1.7133	-

R^2 = 0.4589
Degrees of Freedom = 30

Table 11. Salmonellosis
(Dependent Variable - Rate of Salmonellosis/1,000 Population)

Independent Variable	Estimated Coefficient	T-Value	β-Coefficient
LSPCD	0.0176	1.2006	.235
LPPM	0.0259	1.9237	.474
LSLPC	-0.0032	-0.0509	-.013
LPZE	-0.2060	-0.8796	-.743
LPPY	-0.1589	-1.4422	-.314
LURB	0.0092	1.4525	.299
T_1	0.8868	1.0723	-
T_2	0.9084	1.0884	-
T_3	0.9823	1.1869	-

R^2 = .3235
Degrees of freedom = 30

Table 12. Variables Used in the Regression Analysis*

RHEP - Rate of reported infectious hepatitis/1,000 population
RSHI - Rate of reported shigellosis/1,000 population
RSAL - Rate of reported salmonellosis/1,000 population
SPCD - Water quality variable (plate count of bacterial population for samples of 1 ml taken from the distribution system)
PPM - Population per square mile
PFTF - Proportion of population in the age group five to twenty-five
PZE - Proportion of population in the age grioup zero to eighteen
PPY - Per capita income
SLPC - Per capita state and local government health expenditures
PPS - Percent population served by a community water supply
URB - Percent urban population

* L in front of a variable name denotes logarithm

tends to go up with an increase in these variables. Also as hypothesized, the coefficient for percent population served (PPS) is negative for both diseases; however, the public health expenditure variable (SLPC) has a positive coefficient in one case and is negative in the other. The age variable (percent population five to twenty-five, PFTF) is also positive indicating that as the percent in this age group increases, rate of hepatitis increases. In the hepatitis equation, the hypotheses that the coefficients of SPCD and income are zero are rejected at the 90 percent level. The time correction variables are also significant at least at the 90 percent level. In the shigellosis equation, the water quality variable is significant at the 90 percent level, the age variable at the 97 percent level, and the time variable also significant above the 95 percent level.

In the salmonellosis equation slightly different variables were used. For age, percent population zero to eighteen was used because better results were obtained. Instead of a percent population served variable, a percent population in urban areas variable was used. These two variables are highly correlated, i.e. a high percent population served indicates an urban area and conversely. Using the PPS variable, a positive coefficient was obtained in the regression; the interpretation that an increase in treated water increased disease made no sense. It was felt that because of the high correlation this variable was in fact reflecting the influence of the degree of urbanness on salmonellosis so the urban variable was used instead. (The percent population urban and density variable are not highly correlated so both could be used in the regression.) All variables are in logarithmic form. In the salmonellosis equation, the water quality variable was significant at the 85 percent level and the income variable and urban variable at the 90 percent level.

β-Coefficients in Tables 9, 10 and 11 indicate the relative importance of variables in the regressions. In the hepatitis equation, the age variable is most important followed by water quality and then income; unit changes in the other variables have little effect on disease rates compared to these three. In the shigellosis equation, water quality was most important followed by age, density and income (in that order), and again the other variables have little importance. In the salmonellosis equation, the age variable was most important, followed by density and income. The β-coefficient for water quality variable was largest in the shigellosis equation and least in the salmonellosis equation.

(ii) Prediction

The regression equations based on the sample of 39 counties can be used to predict U.S. reported rate of disease; a comparison of the predictions with actual reported U.S. rates gives an indication of how valid the regression equations are in predicting disease.

Table 13 gives mean values of the independent variables for the sample

Table 13. Comparison of Sample, U.S., and Predicted Value

Variable	39 Counties Average Value	Actual U.S. Value	Predicted U.S. Value
SPCD	305.351	305.35*	-
PPM	1212.03	56.9	-
SLPC	5.531	5.48	-
PFTF	.3827	.37	-
PPY	2991.56	3119	-
PPS	.7431	.75	-
PZE	.3645	.35	-
URB	.5652	.69	-
T_1	.2564	.256*	-
T_2	.3076	.307*	-
T_3	.4358	.436*	-
RHEP	.2292	.240	.2001
RSHIG	.0402	.091	.0314
RSAL	.0742	.059	.0706

* U.S. Value unknown, value taken from CWSS sample

and for the U.S. for 1969 and the predicted U.S. reported disease rates for 1969. The mean values for the U.S. are used to predict U.S. rates from the regression equations. (In the case of salmonellosis, the logarithms of the variables are used in the prediction.) From CDC (1971), the actual rates per thousand in 1969 are 0.240 for hepatitis, 0.091 for shigellosis and 0.059 for salmonellosis. The predictions are quite close in the cases of hepatitis and salmonellosis but not as close in the shigellosis case. Again, the purpose of these predictions is to test reliability of the models; the results indicate that the hepatitis equation and salmonellosis equation have validity as models of reported disease rates.

The error of the prediction depends on two factors, the amount of error in the regression equation in estimating rates for the sample and by how much the characteristics of the U.S. differ from the sample characteristics. U.S. density is much smaller than density for the sample and income is a little higher in the U.S. Apart from the Community Water Supply Study, water quality in the United States as a whole is not really known. The water quality value of 305.35 used in the U.S. prediction is taken from the county average; this is the best estimate of U.S. water quality available. County averages for the time correction factors are also used in the U.S. prediction. U.S. predicted rates are quite close to the average of county rates. Differences in the predicted and actual U.S. values are due mainly to the differences in the sample disease rates and U.S. rates; particularly in the case of shigellosis, the sample rate is much lower.

A major use for the disease regressions is to estimate what part of reported disease is actually due to water quality. Using the predicted rates for the U.S. and assuming that reducing SPCD from 305.35 to zero eliminates all water-caused disease, the percent of reported disease due to water can be calculated using (2) in the case of hepatitis and shigellosis as

$$\frac{\Delta RHEP}{RHEP} = \frac{(305.35)(0.00016)}{.2001} = .244$$

$$\frac{\Delta RSHI}{RSHI} = \frac{(305.35)(.00011)}{.0314} = 1.07$$

and using (5) in the case of salmonellosis as

$$\frac{\Delta RSAL}{RSAL} = \frac{.0176}{.076}\,\frac{\Delta SPCD}{SPCD} = .249$$

(Note: That $\frac{\Delta R}{R} = \frac{\Delta(\#\text{ cases})/pop}{\#\text{ cases}/pop} = \frac{\Delta(\#\text{ cases})}{\#\text{ cases}}$ so the percent change in the rate is the same as the percent change in the number of cases.)

Applying these percentages, 24.4, 100 and 24.9 percent, to the reported cases of the three diseases for the U.S. yields estimates for 1969 of 11,814 out of 48,416 reported cases of hepatitis, all 18,419 reported cases of shigellosis, and 2,975 out of 11,946 reported cases of salmonellosis as actually due to water quality among the reported cases. Table 14 compares these estimates with the reported waterborne outbreaks of these diseases and shows that among the total reported cases, the predicted water related disease is much higher than that with source identified positively as being due to water.

Table 14. Reported Cases Predicted to be Water Related, U.S. 1969

	Total Reported Cases of Disease**	Cases with Source Identified as Water*	Cases Predicted Related to Water
Hepatitis	48,416	111	11,814
Shigellosis	18,419	90	18,419
Salmonellosis	11,946	131	2,975

* Craun & McCabe (1973)
** Center for Disease Control (1972)

C. Production Function for Water Quality

In order to compute benefits of a specific program, the effects of the program on water quality must be determined; the reduction in reported disease due to the program can then be determined using the disease relations given above.

The determination of water quality as related to surveillance, source, raw water quality, treatment and operator characteristics was limited in this study by the data available in the Community Water Supply Study; the data was limited both in the type of questions asked and the number of responses. Hence, the following discussion mainly demonstrates a method of analysis which could be applied with better results if better data were available.

Information on treatment processes (clarification, disinfection, etc.) was available only as to whether it was adequate or inadequate and many questions lacked response to these questions. Information on processes related to quality of the distribution system (booster chlorination, residual chlorine, pressure, etc.) was also not given for many of the systems. Only the treatment processes included in the questionnaire with sufficient response could be used in the production equation, eg. the effect of ozone on water quality cannot be included. Likewise, the effect of short courses and additional training for operators on water quality could not be determined due to lack of responses to the CWSS questionnaire. The only education variable available was the amount of formal training an operator had and this may not be a very good indicator of ability. A variable indicating position of the main operator (part-time or full-time) was also included. (The position data on the CWSS tape seems somewhat contradictory.) Surveillance data (including the number of months samples weren't taken, the number of months samples exceeded quality standards, and the number of months insufficient samples were taken) for each system was combined into a single index to indicate quality of surveillance (with surveillance increasing from 1 to 37). A population variable was also used to account for system size; although a larger system would tend to have better treatment at the plant, the distribution system would also be longer which would give more chance for contamination in the system to occur. Also, more education would be required to operate a larger system. Thus, the variable POP is used to adjust for differences in characteristics of systems which are due to population.

Water quality in the distribution system (SPCD) was measured similarly to that described above for the disease equations (# bacteria/ml). The main difference is that here it is an average over samples in a given distribution system while for the disease equations it represented average quality in a county and there may be several systems in a county. Raw water quality (SPCS) is also measured by the Standard Plate Count (# bacteria/ml) and is also an average for each system. More details on how all these variables were measured is in the appendix. Table 15 gives the variables and their mean values for the 870 systems.

All variables which are indexes are in increasing order of quality starting from 1 (SURV, CLAR, DIS, ED, POS). A regression relation between water quality in the distribution system (SPCD) and the vari-

Table 15. Characteristics of 870 Water System Sample*

Variable	Mean Value
SPCD (bacteria/ml in distribution system)	389.52
LSPCD	3.34
SPCR (bacteria/ml in raw water)	340.40
LSPCR	0.95
RC (residual chlorine in ppm x 100)	13.56
SURV (surveillance; 1 to 37 in order of increasing quality)	17.18
MIX	.08
GRND (zero-one variables indicating source)	.71
SURF	.21
CLAR (clarification adequacy; 1. 2. 3 in order of increasing quality)	1.97
DIS (disinfection adequacy; 1, 2, 3)	1.79
ED (education level of operator; 1, 2, 3,4)	1.97
POS (presence of full or partial operator; 1, 2, 3, in order of increasing time)	1.88
POP (population	11392.
T_1	.40
T_2 (time correction factors)	.13
T_3	.47

* In the case of logarithmic variables, mean values are the average of the logarithmic, or the geometric mean.

See Appendix for further explanation of variables.

ables described earlier was estimated. A logarithmic form for SPCD and some of the independent variables gave the best fit to the data. Table 16 gives the coefficients for this regression (the letter "L" before a variable indicates a logarithmic form, otherwise the variable is in linear form). The variables for source (MIX, GRND) are zero to one variables with surface water indicated by zero for MIX and GRND; these variables serve only to shift the intercept. (The source of the water of course explains the quality of raw water but this is not the point of this equation.) Time correction factors as explained above are also included here.

The regression is thus of the form

$$\log \text{SPCD} = \Sigma_i \alpha_i \log X_i + \Sigma_j \alpha_j X_j + \theta \qquad (6)$$

or

$$\text{SPCD} = e\theta \, \Pi j e \alpha_j X_j \, \Pi_i X_i \alpha_i \qquad (7)$$

Table 16. Regression Equation for LSPCD

Variable	Coefficient	Standard Error	T-Value	β-Coefficient
LSPCR	0.1153	0.0189	6.095	.195
MIX	-0.4570	0.4579	-0.998	-.035
GRND	-0.7132	0.3334	-2.139	-.096
LSURV	-0.1471	0.0695	-2.117	-.069
LCLAR	0.5648	0.4210	1.342	.055
LDIS	-0.0508	0.2427	-0.209	-.006
LED	0.2701	0.2530	1.068	.037
LPOS	-0,3613	0.4647	-0.781	-.027
RC	-0.0284	0.0032	-8.630	-.300
POP	.00000299	.00000206	1,451	.047
T_1	2.143	0.2302	9.312	-
T_2	2,204	0,3542	6,223	-
T_3	3,009	0,382	7,870	-

R^2 = .1977; degrees of freedom = 857

where θ is the intercept, the subscripts "j" indicate linear variables, and "i" indicate variables used in logarithmic form. Percent changes in water quality due to each type of variable are predicted as follows:

$$\frac{\Delta SPCD}{SPCD} = \alpha_j \Delta X_j \qquad (8)$$

for the linear variables and

$$\frac{\Delta SPCD}{SPCD} = \alpha_i \frac{\Delta X_i}{X_i} \qquad (9)$$

for the logarithmic variables, where SPCD is water quality corresponding to given values of the independent variables, ΔX_j is the change in a linear variable with all other variables constant, and ΔX_i is a change in a logarithmic variable with all other variables constant. If several variables are changed simultaneously, then the corresponding changes are summed to get the total change in SPCD. Use of the regression for computing changes in SPCD in explained in more detail below.

(i) Hypothesis Testing and Importance of Variables

As expected, the sign of the coefficient for SPCR was positive indicating that lower quality raw water is associated with lower quality water in a distribution system; the variable is also significant at the 99 percent level. The disinfection variable has a negative coefficient indicating that an increase in disinfection is associated with a decrease in SPCD; however, the t-value shows the coefficient is not significantly different from zero. Likewise the coefficient for opera-

tor position has the hypothesized sign but is not significant. Residual chlorine has a negative coefficient showing that an increase decreases SPCD and is significant at the 99 percent level. The coefficient of the population variable is significant at the 90 percent level. The signs of the coefficients for ED and CLAR seem to be counterintuitive (an increase in these variables is associated with worse quality water) but are not significant. An explanation in the case of CLAR is the large number of no responses on the questionnaire; there were only 166 responses out of 870 systems. The education variable used (degree of formal training) is probably a poor indicator of quality of operator.

The β coefficients can again be used to compare importance of variables in the regression. Residual chlorine is most important followed by raw water quality, source of water and population. Operator position and disinfection are least important.

Of course, successful maintenance of residual chlorine cannot be accomplished without disinfection, clarification and presence of an operator. These results are not meant to suggest that these ingredients of good water quality are not important. The regression coefficients do not reflect the interdependencies in these processes, merely the effect of each process considered separately. Hopefully, with better data a better regression equation can be estimated.

(ii) Prediction

The main use of the regression equation is to predict effects of changes in variables on water quality. There is little other prediction that could be done at this point. If average values of the water treatment variables were known for the U.S., the equation could also be used to predict average U.S. water quality; however, the average values in Table 15 represent the best U.S. estimates available for these variables.

To predict effect of changes in variables on average SPCD, equations (8) and (9) are used. For example, if residual chlorine is changed from 0.13 to 0.30 then from (8) the percent change in SPCD is given by

$$\frac{\Delta SPCD}{SPCD} = -.028 \quad \Delta RC = -.028\ (16.44) = -.46$$

or a 46 percent decrease. If POS is changed by 7 percent, then from (9)

$$\frac{\Delta SPCD}{SPCD} = -.361\ \frac{\Delta POS}{POS} = -.361\ (.07) = -.025$$

or a 2.5 percent decrease.

D. Estimating Disease Reduction Due to Specific Water Programs

Now that the two basic types of relations (disease and water quality) have been developed, their use together in analyzing programs can be demonstrated. For illustration purposes, a specific program package described earlier will be analyzed. The program package to be analyzed is:

1. To bring all community water supplies into compliance with surveillance standards
2. To ensure that all systems operate adequately with respect to disinfection
3. To require that all systems serving 0 to 1000 people have a part-time operator and a full-time operator for systems serving over 1000 people. (It is assumed that maintenance of a residual chlorine requires at least a part-time operator.)
4. To have all distribution systems maintain a residual chlorine of 0.3 to 0.5 ppm

The parameters of this program package (surveillance, treatment, residual chlorine and operator characteristics) correspond to those in the water quality production relation. Thus by transforming these program parameters into measures used in this study, the effects of this program package on water quality can be analyzed.

Knowing the effect on water quality, effects on disease can then be estimated using the disease regressions reported earlier. Tables 17 and 18 summarize the results on water quality and disease; details of computations for the tables will be given.

Table 17. Improvement Programs and Their Effect on Average Drinking Water Quality

Program Parameter	Proposed Average Level***	% Change over present mean Level	% Change in SPCD
I. Surveillance	30.74	78%	-11.5
Disinfection	2.213	25%	- 1.3
Position	2.54	25%	-12.6
Residual Chlorine	.30	130%	-45.9
		Total % Reduction in SPCD	71.3%
II. Surveillance	30.74	78%	-11.5
Disinfection	2.213	25%	- 1.3
Position	2.54	35%	12.6
Residual Chlorine	.40	190%	73.9
		Total % Reduction in SPCD	99.3%

Program Parameters		Proposed Average Level***	% Change over** Present Mean Level	% Change in SPCD
III.	Surveillance	30.74	78%	-11.5
	Disinfection	2.213	25%	- 1.3
	Position	2.54	35%	12.6
	Residual Chlorine	.50	280%	101.9
			Total % Reduction in SPCD	100%

*Clearly SPC cannot be reduced beyond 100 percent; the excess over 100 percent just represents a safety margin of residual chlorine.
**Mean level is the level reported in the CWSS study.
***See definitions of parameters in Table 15 and Appendix.

Table 18. Reduction in Average Reported Disease Rates Due to Improvements in Drinking Water Quality

Program	Disease	Percent Change in Rate of Reported Disease*
I	Hepatitis	17.39%
	Shigellosis	76.29%
	Salmonellosis	17.75%
II	Hepatitis	24.22%
	Shigellosis	100 %
	Salmonellosis	24.72%
III		
III	Hepatitis	24.40%
	Shigellosis	100 %
	Salmonellosis	24.90%

*Same as percent reduction in disease cases since

$$\frac{\Delta \text{Rate}}{\text{Rate}} = \frac{\Delta(\#\text{ cases})/\text{ POP}}{\#\text{ Cases/POP}} = \frac{\Delta(\#\text{ cases})}{\#\text{ cases}}$$

(i) Surveillance

Using the 870 sample data as being most representative of unknown U.S. conditions, systems now complying with standards because they serve interstate carriers had an average surveillance score of SURV = 30.74 while in the CWSS study the overall (including non-interstate carrier systems) average score was 17.2. Assuming that compliance by all systems means they will then act similarly to interstate carrier systems the new average SURV corresponding to point one of our program would be 30.74. The change in average SURV due to the program is thus

$$\Delta \frac{\text{SURV}}{\text{SURV}} = \frac{30.7 - 17.2}{17.2} = \frac{13.5}{17.2} = .78$$

or a 78% increase.

Using equation (9), the change in average SPCD is then

$$\frac{\Delta SPCD}{SPCD} = -0.147\ (.78) = -.115$$

(ii) Disinfection

From the data on disinfection, the average score was 1.79 (this was obtained by assigning a score of 2 to ground water sources with "blank" disinfection information). Assigning a score of 3 for disinfection to all except ground water systems and using the distribution of sources in the sample (21.3% surface and mixed, 78.7% ground) yields a new average disinfection score of

$$DIS = 3(.213) + 2(.787) = 2.213$$

for a percent change of 25%

The resulting change in average SPCD is then obtained from equation (9) as

$$\frac{\Delta SPCD}{SPCD} = -.051\ (.25) = -.013$$

(iii) Operator Position

For the 870 system, the average operator position is POS=1.88 indicating that most systems have at most an occasional or part-time operator. Assuming that the distribution of the CWSS systems between serving under and over 1000 people is representative (54% over 1000, 46% under 1000), under the program above the new POS variable would be

$$POS = 2(.46) + 3(.54) = 2.54$$

giving

$$\Delta POS = 2.56 - 1.88 = .66$$

and a percent change of 35%.

The change in average SPCD for this change in average Position is then

$$\frac{\Delta SCE}{SPCD} = -.361\ (.35) = -.126$$

(iv) Residual Chlorine

For the 870 system data, the average r idual chlorine was 0.13 ppm. Under a program with RC = 30 (residual chlorine at 0.30 ppm)

$$\Delta RC = 30 - 13.6 = 17.4$$

With RC = 40

$$\Delta RC = 40 - 13.6 = 26.4$$

and a RC = 50 program yields

$$\Delta RC = 50 - 13.6 = 36.4$$

The respective changes in average SPCD due to these respective changes in RC calculated from equation (8) are:

$$\frac{\Delta SPCD}{SPCD} = -.028\ (17.4) = -.459$$

$$\frac{\Delta SPCD}{SPCD} = -.028\ (26.4) = -.739$$

$$\frac{\Delta SPCD}{SPCD} = -.028\ (36.4) = -1.019$$

(v) Disease Calculations

Using the percent reductions in SPCD ($\frac{\Delta SPCD}{SPCD}$) given in Table 17 for the various programs, the estimated present average value of SPCD of 305.35 and the disease regression relations presented above, the reductions in disease rates are obtained from the equation

$$\Delta Rate = \frac{\alpha, \frac{\Delta SPCD}{SPCD}\ 305.35}{Rate}$$

derived from (2) in the case of hepatitus and shigellosis (where α, is the coefficient of SPCD in the disease regression). In the case of salmonellosis the equation used is

$$\Delta Rate = \frac{\alpha, \frac{\Delta SPCD}{SPCD}}{Rate}$$

derived from 5.

For program I, the calculations are

$$\frac{\Delta RHEP}{RHEP} = \frac{(.00016)\ (.713)\ (305.35)}{.2001} = .174 \text{ or } 17.4\% \text{ reduction}$$

$$\frac{\Delta RSHI}{RSHI} = \frac{(.00011)\ (.713)\ (305.35)}{.0314} = .764 \text{ or } 76.4\% \text{ reduction}$$

$$\frac{\Delta RSAL}{RSAL} = \frac{(.0176)\ (.713)}{.0706} = .178 \text{ or } 17.8\% \text{ reduction}$$

For program II

$$\frac{\Delta RHEP}{RHEP} = \frac{(.00016)\ (.993)\ (305.35)}{.2001} = .242 \text{ or } 24.2\% \text{ reduction}$$

$$\frac{\Delta RSHI}{RSHI} = \frac{(.00011)\ (.993)\ (305.35)}{.0314} = 1.06 \text{ or } 100\% \text{ reduction}$$

$$\frac{\Delta RSAL}{RSAL} = \frac{(.0176)\ (.993)}{.0706} = .247 \text{ or } 24.7\% \text{ reduction}$$

For program III

$$\frac{\Delta RHEP}{RHEP} = \frac{(.00016)\ (1)\ (305.35)}{.2001} = .244 \text{ or } 24.4\% \text{ reduction}$$

$$\frac{\Delta RSHI}{RSHI} = \frac{(.00011)\ (1)\ (305.35)}{.0314} = 1.07 \text{ or } 100\% \text{ reduction}$$

$$\frac{\Delta RSAL}{RSAL} = \frac{(.0176)\ (1)}{.0742} = 2.49 \text{ or } 24.9\% \text{ reduction}$$

These percent reductions are presented in Table 18. They may be converted to reduction in number of reported disease cases by multiplying the percents by the number of reported disease cases. (Recall that the percent change in the rate is the same as the percent change in the number of cases.) Reductions in reported disease cases presented in Section VII.

To obtain estimates of actual disease reduction from reported disease reduction for the various programs, a multiplier would have to be applied to reported disease. However, it may be assumed that the percent reductions in actual disease would be the same as the percent reductions in reported disease.

Thus, varying the residual chlorine requirement and using the other program requirements, different levels of reduction in reported disease are obtained. The RC = 40 requirement nearly eliminates all water caused disease while the RC = 50 requirement continually does so. The

above percent reductions for RC = 50 (residual chlorine = 0.50) will be used to estimate benefits in Section VII. This completes the demonstration of the use of the two types of equations in analyzing water quality programs. As demonstrated here, the two types of equations are a powerful tool in the analysis of water related disease and how such diseases are affected by water programs.

E. Final Observations

We have developed and demonstrated a method of analyzing effects of raw water quality improvement programs. A specific improvement program was analyzed here for illustration purposes. Though the program analyzed is a reasonable one and consistent with EPA suggestions for improvement, other types of programs could certainly also be considered and analyzed using this methodology. However, the results and applicability of this study are limited by the quality and nature of questions and responses to the CWSS. Given a better survey (and better estimates of U.S. water quality) more accurate predictions of disease reductions could have been made. Also, given more information by system on treatment and operator characteristics, other types of programs could have been analyzed.

It should be pointed out that the reduction in disease estimates obtained here refer to the year 1969 because the relationships were measured using data on disease, and water quality for 1969; the sociological data used was as close to 1969 as possible (for 1967 in the case of health expenditures and 1970 for social characteristics; it was felt the relationship of these variables across counties would be similar to that in 1969). Provided it is available, predictions for years other than 1969 can be made using water quality and sociological data. These predictions will be valid as long as there are no structural changes such as in methods of reporting disease.

In addition to analyzing program effectiveness, the disease equations here give a way of estimating how much of reported disease is actually related to water, either directly or indirectly. As was shown, this figure is much higher than that disease with source identified as to water. It must be emphasized that the estimates here are in terms of reported disease, i.e. there is another aspect of disease, (that which is unreported) that is not covered in this chapter. It can be assumed that the reasons for not reporting disease are the same regardless whether due to water or other causes (i.e. availability of doctors, severity of disease, etc.). Given an estimate of the ratio of unreported to total disease (which could range from 10 to 100 according to Askeraff et al. [1970] and Barker et al. [1969]), this ratio could be used to extrapolate the predictions of reported disease due to water to actual disease due to water and actual disease reductions due to a program.

Once disease reduction estimates corresponding to a program have been made, benefits of the program may be computed by multiplying disease reduction by cost per incidence of disease. This benefit estimate may then be compared to costs of a program to obtain a benefit/cost analysis of the program. Since cost in a benefit/cost analysis refers to a specific program, it makes no sense to make benefit/cost comparisons unless the benefit measurement also refers to the same program. The importance of the method presented here is that it enables such program-specific comparisons to be made. The next chapter deals with estimates of cost of the program analyzed here and the final chapter deals with benefit/cost comparisons for this program.

SECTION VI
COST OF DRINKING WATER QUALITY IMPROVEMENT PROGRAM

The purpose of this section is to present a methodology for evaluating the costs of various water quality improvement measures for use in studies of the relation of the resulting economic benefits to their costs. Costs and other figures used are illustrative over-all data, and are based on data presently readily available. As further studies on benefit/cost relationships proceed, better costs figures should be developed, and the effects of economy of scale should be applied to the "population" of water systems of various sizes.

Another purpose of this section is the identification of areas of uncertainty, to which more concrete effort can be directed, in order to obtain information that will eliminate the uncertainties. In addition, where possible, judgement estimates have been made of resulting benefits in reduction of exposure to bacterial and viral pathogens. In one case the economic benefit was calculated directly, and in another, it was approximated.

Population Served, Personnel Required and Basic Cost Data

A current study by Thursby, et. al. (1974) states that the number of community systems is about 49,000. In addition there are "other" systems serving the public which total 230,000. An EPA inventory yielded values of 40,000 public water supply systems and 200,000 "other" systems (Anon. EPA, 1973, 1973a). The 40,000 community water supply systems serve about 130 million persons (McCall, 1974), thus the average size of these is about 3,250 persons per system. The 200,000 systems serve about 50 million people, and their average size is thus about 250 persons per system. A 1970 study (Hudson and Rodrigues) showed that it took an average of one man to operate a U.S. community water utility serving about 250 or less people, while for larger community utilities there was an average of 1.1 employees per 1,000 served.

Applying the 1.1 per thousand figure to the 130 million people served by community systems, and applying the one employee per system figure to the 200,000 other systems yields the following estimates:

Type of System	Number of Employees
Community	143,000
Other	200,000
TOTAL	343,000

The estimate of 200,000 full-time employees is probably higher than any potential reality, since the small "other" systems operation is nearly always a part-time activity. Assuming it to be a one-third time job, the total number of water system employees may be taken to be about 200,000.

Community water supply systems have had an estimated construction value of $70 billion, or $540 per person served. The replacement cost of these systems was estimated (Anon. 1974) at $125 billion or nearly $1,000 per capita served. The "other" water supply systems, ordinarily built under less rigorous design criteria and circumstances, may be estimated to have cost $300 per capita, for a total value of about $15 billion. Annual billings of community water systems are $2.5 billion, or nearly $20 per person served (McCall, 1974). On this basis, annual charges in the "other" systems may be estimated to be approximately one billion dollars.

Water Utility Salaries and Training Costs

If we assume a proper annual compensation rate of $10,000 per employee for the 200,000 employees, we arrive at an annual payroll of $2.0 billion per year. The American Water Works Association recommends that there be budgeted for training a minimum of 5 percent or about $100 million per year. This amounts to 55 cents per capita served annually. Present levels of funding for training are believed to be generally lower than the 5 percent figure.

Moreover, actual salary rates in water utilities are presently lower than the $10,000 figure and very likely average about $7,000 per employee annually, for a total of approximately $1.40 billion for community and "other" systems. This would leave an estimated $600 million needed annually to overcome inadequate salaries in the industry. The current salary estimate appears consistent with the estimated annual charge for water service of $2.5 billion per year for community and other systems. It does not appear that there are surplus funds presently available to provide the estimated $600 million deficit.

Without rectifying salary deficiencies, training will be relatively ineffective, for there will continue to be a high turnover of personnel, particularly of those who do receive training. This is borne out by much prior experience in the U.S. and abroad. Training a person while compensating him meagerly provides a strong incentive for him to move to a better-paying job elsewhere, using the skills acquired under the training. Such people usually move completely out of the water utility industry.

A 1968 survey by Dyer (1974) showed that the weighted average salary of water utility managers was $10,540 annually. The average cost of each community utility has been nearly $1.5 million, a weighty responsibility for a manager compensated less than an engineer-in-training.

Based on a total population served of 180 million persons, the $600 million salary deficiency amounts to about $3.33 per capita served. It may not be reasonable to expect this deficiency to be corrected in a single year. If it is assumed to be phased in over a ten-year period,

to provide a new base for budgeting, the resulting annual figure of $60 million looks reasonable. Adding in the $0.55 per capita cost of training, the per capita cost comes up to $3.88 annually at the end of the tenth year.

Monitoring and Surveillance Costs

The cost of state monitoring and enforcement has been worked out in a study by the National Sanitation Foundation to be $146 million annually (Thursby, et. al., 1974). Funding of state monitoring programs is reported to be currently at about 20 percent of that level. The Safe Drinking Water Act, would provide $15 million in Fy 76 and $25 million in Fy 77 for the federal grants in the drinking water quality field. This is still $98 million less than required annually.

Surveillance, Training and Chlorination

Even without major improvements in water treatment facilities, a great deal can be done at modest costs through more effective surveillance and training. These two are closely tied together: Surveillance can be used to create an impulse toward training; and training creates a better recognition of the value of usefulness of surveillance. Surveillance without training activities tends to lead the operators into feeling that they are functioning in a police state. Without training, surveillance is resented and is less effective than it should be.

If the combination of surveillance and training was used to identify the more hazardous sources, and if chlorination was instituted only at those sites, it is estimated, on the basis of the author's experience, that waterborne viral disease could be brought down 30 percent. It appears reasonable to assume that less than 10 percent of the systems have potentially hazardous or "high risk" sources. Such high-risk sources might include, for example, untreated surface supplies, or untreated well water in limestone terranes thinly covered by unconsolidated earth materials. The 1969 community water supply study identified 19 percent of the systems as being of dubious quality. These are readily identifiable by sampling and analysis and by sanitary surveys.

Provision of simple chlorination, the major deficiency in small systems throughout the United States, would cost approximately $5,000 for installation on a system serving 1,000 people. Costs at smaller utilities would be lower. This is a capital cost of about $5 per person which at the interest rates prevailing over the past decades, could be amortized and operated at a cost of about 75 cents per capita annually. Extending the $5 per capita figure to the remaining population receiving unchlorinated water, estimated at about 45 million people, yields a total capital expenditure for simple chlorination of $225 million and an annual expenditure of about $22.5 million. It is estimated that the full initiation of effective chlorination of all drinking water systems would reduce shigellosis, salmonellosis, and other bacillary waterborne diseases by 95 percent or more, and viral disease would probably be reduced by an

additional 10 percent beyond the reduction attained by chlorination of only the more hazardous sites. If we use 10 percent as an estimate of the number of possible high-risk sites, the $250 million figure could be cut $25 million. This would probably be done as an opening step in the program, with the objective of moving toward chlorination of all supplies in time, the public health objective that is generally agreed upon amongest water quality surveillance specialists.

Softening and Clarification Plants

A further step in bettering water quality treatment is clarification or softening plants. Clarification plants are used to accomplish improvements in the physical and biological characteristics of the water. They usually involve chemical treatment, mixing, flocculation, settling, chlorination and filtration; and the combination of these processes normally produces virtually 100 percent removal of algae and bacteria, and about 98 percent removal of viruses. Softening plants accomplish all that clarification plants attain, but, in addition, probably secure even better virus removal, and they improve the mineral quality of the water by reducing objectionably high hardness levels to acceptable levels. Softening also reduces most of the heavy metals to levels below those of health concerns.

Figure 2 shows the relationship of plant capacity to new plant construction costs for clarification plants of two characterizations: those of conventional design; and plants of high rate design. These curves also apply reasonably well to softening plants although cost of operation of such plants is higher than that for clarification plants.

If we assume that the plants need to be built to produce at peak capacity rates of 200 gallons per capita daily, a high-rate 10 mgd plant could serve a population of 50,000 people and be built at a per capita construction cost of $40. On the same basis, a 300 mgd high-rate plant serving 1.5 million people would cost about $10 per capita served because of economy of scale. Most large cities which need such treatment already have plants, hence a $33 per capita estimate average cost seems reasonable. If we assume that such plant facilities are needed for a population of 50 million people, the construction cost of such facilities could be estimated at $1.65 billion. It is assumed that such facilities could be operated and amortized, which yields an annual expense of $165 million per year. This, in conjunction with an adequate training, surveillance and chlorination program, could wipe out nearly all water-transmitted diseases, including almost all water-transmitted hepatitis.

Exotic Organic Substances and Heavy Metals

Viruses are known to be well removed by absorption (Cookson and North, 1967). Concentrations of many of the heavy metals are greatly reduced by passing the water through activated carbon beds (Sigworth and Smith, 1972). There are major side benefits, in that the carbon treatment virtually eliminates

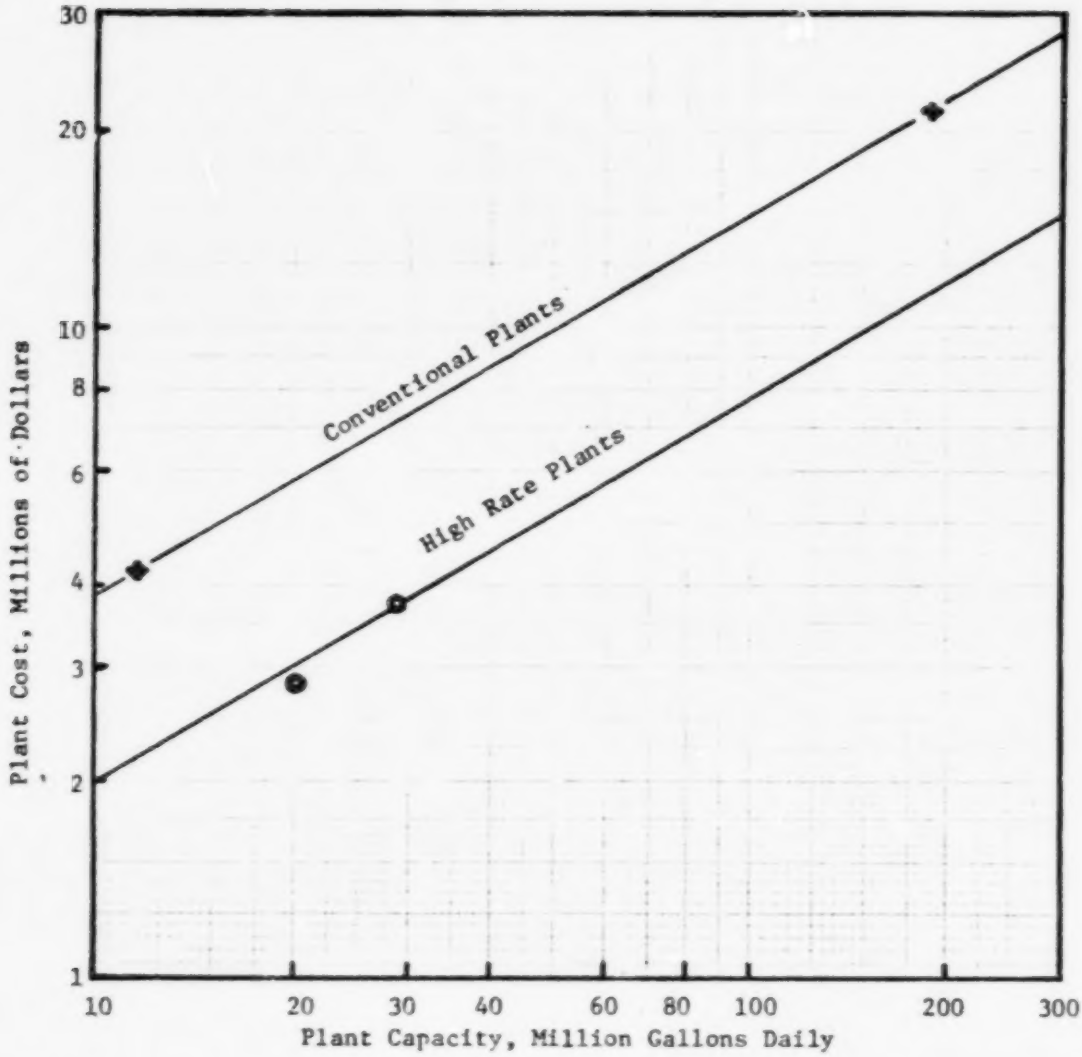


Figure 2. Cost of Construction of Water Treatment Plants, 1974 Dollars

tastes and odors, and reduces color. Assuming an average use of 150 gpcd, annual per capita costs of carbon treatment, including amortization and operation, if built as a separate unit operation, based on Figure 3 would run from $4.30 in a 100 mgd system, to $13.80 in a 1 mgd system (Monti and Silberman, 1974). Capital costs over a similar range, again based on Figure 3, would range from $50 to $160 per capita served. These annual per capita costs include amortization and operating costs of carbon handling and regenerating equipment. Since these costs were developed for advanced waste treatment, they are probably considerably on the high side.

Most of the community water systems in which active carbon treatment would be needed already have water treatment plants. When activated carbon is substituted for anthracite on sand filter media in filters, the additional cost of providing carbon treatment is much less (Monti and Silberman, 1974). A further benefit lies in the potential for plant up-rating which results from use of dual media beds; up-rating of plants usually requires other expenditures for increasing hydraulic capacity, etc., so that simply changing media is not the total solution.

Such carbon treatment costs about $1.25 per cubic foot, and has an operating life averaging about three years. Assuming a 2-foot thick bed operated at 2 gpm/sf, and using a per capita quantity of 150 gpd of water, one square foot of filter bed would serve about 19 people. This works out to an annual per capita cost of $0.04. The per capita capital cost is $0.12. No regeneration is assumed: the carbon is discarded after three years.

Based on actual experiences at Mt. Clemens, Michigan, annual per cpaita costs (assuming 2.5 persons per family) was $.24, and capital cost of initial installation was $.72 per capita (Hansen, 1972). If it is assumed that carbon bed treatment would be desirable for the systems supplying water to 50 million people, on the basis of the Mt. Clemens experience, a total annual cost of $12 million results. The initial investment is $36 million.

As yet, there are few data on the extent or severity of illness caused by exotic organics, so that health-related benefit/cost evaluations cannot be attempted. Heavy metal concentrations in water are believed, almost everywhere, to be below the threshold of causation of illness. Activated carbon treatment for exotic organics in known not to be universally successful. In the circumstances we have to assume that treatments will be developed which will be effective at costs comparable to those of carbon treatment.

One way to evaluate the worth of taste and odor elimination would be to assume that the universal institution of such treatment might eliminate at least half, and perhaps all, of the bottled water industry in the U.S. That industry currently produces revenues in the order of $100 million per year. Assuming that the benefit so assayed is $50 million annually, on the basis of the Mt. Clemens experience, this illustrative example would have a benefit/cost ratio of 4:1.

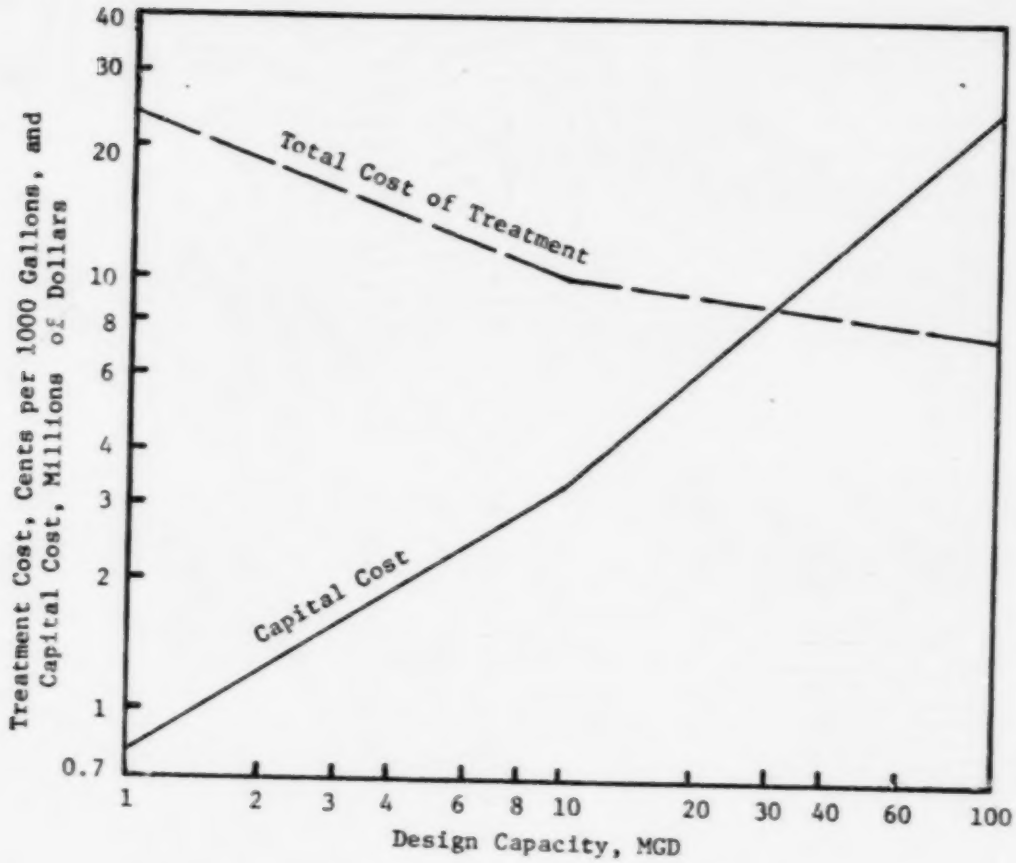


Figure 3. Cost of Carbon Absorption Treatment. After Monti andSilberman. 15 Minute Detention.

Corrosion Control

Another measure which should be considered is the stabliization of water to make it non-corrosive and non-encrustant. Some waters are naturally stable, and others are made stable by chemical treatment, but many waters are distributed in a chemically unstable condition, usually on the corrosive, rather than the encrustant side. Improvement of unstable waters should be part of any water quality improvement program. Generally, only the larger water utilities do this effectively, and not all of them do. Boston and New York, for instance, distribute corrosive waters.

Investment in community water distribution systems comprises about 60 percent of the capital cost of water utilities. Total construction cost of community systems has been $70 billion, or $540 per capita. Sixty percent of the total is $42 billion. It has been determined that the water distributed by nearly half of the 100 largest cities in the U.S. is unstable (see table below).

Table 19. Stability of Water Distributed by 100 Largest Cities in U.S. (Based on Langelier Index)

Stable	49	
Corrosive	46	
Questionable	5	(0< Index< 0.1)

Even with stable waters, distribution capacity of new piping declines with age at a rate of one percent per year. With unstable waters the decline commonly doubles this, producing an excess lost of one percent per year (W.D. Hudson, 1966). Based on the assumptions stated above, it appears that good stabilization treatment could enable avoidance of an economic loss of $210 million annually.

In most instances, corrosion control can be attained by adding lime to the water (an approximation might be 20 mg/l). At a per capita use of 150 gallons daily, stabilization would require about 9 pounds of lime per capita which, assuming a cost of $30 per ton, would price out at 14 cents per capita annually. A safe estimate, allowing for operating costs and amortization of facilities, would be 30 cents per capita annually. Added investment in equipment is estimated at $40 million.

Assuming that one-half of the population served by community water systems is provided with corrosive water, the annual cost of rendering this water stable would be $20 million, which would be the cost of avoiding a $210 million dollar annual loss. Well-run utilities have recognized this problem, and have long ago instituted corrective actions.

This factor is a simple economic one. It may also involve some health benefits. (See Appendix D.) A separate investigation now in progress, is tentatively indicating that the application of corrosion control to those utilities not currently applying such treatment effectively would reduce cardiovascular disease at those locations by 10 per cent.

It carries with it the large intangible benefits of elimination of red water and reduction of quality deterioration in distribution systems, thus making the water far more acceptable and usable to consumers.

Summary

Cost estimates and other estimates contained in this section are summarized in Table 20. It has been assumed that salary up-grading, training, surveillance and limited application of chlorination of high-risk supplies would be executed simultaneously as a first step. The table includes some crude preliminary estimates of reductions in heavy metals. A literature search would lead to much-improved estimates for these values.

The estimates given in the text and in the tables that follow are preliminary, and are only for purposes of illustrative methodology. Some of the estimates are incompletely explained, but, nonetheless, serve the purpose of illustrating approaches to the problem. As work on benefit/cost evaluation progresses beyond the stage represented by this preliminary review, much better, more detailed, estimates should be anticipated.

The estimated costs of the measures required to provide essentially complete protection against health risks, in comparison with present costs, together with estimated disease rate reduction are recapitulated in Table 20, and given in more detail in Table 21.

Table 20. Summary of Estimated Costs of Various Water Quality Programs, and Estimated Reduction in Waterborne Disease

Items	Million Dollars		Resulting Cumulated Reduction in Water-Induced Disease Incidence, Percent*		
	Capital	Annual	Bacillary	Viral	CVD
Present Costs	85,000	3,500	-----	---	---
Cumulative Additional Costs to Accomplish various Quality Levels					
Up-grading Salaries, Training, Surveillance and Limited Chlorination	25	8863	90	30	---
Universal Chlorination	250	885	95	40	---
Clarification and Softening	1,900	1,050	100	98	---
Activated Carbon Treatment	1,936	1,032	100	100	---

Corrosion Control	1,996	1,082	---	---	10**

*Estimates based on author's experience.
**Based on study in progress by Hudson.

In one way, the estimated changes in exposure to various hazards will appear redundant for, under another sector of this study, relations between disease incidence, chlorination and filtration are being evaluated by regression analysis. That analysis will not, however, yield information on effects of softening, carbon bed filtration, nor on exotic organics or heavy metals. The regression study also can be expected to undervalue the capabilities of chlorination and clarification treatments under conditions of higher quality training, surveillance and operation than those presently prevailing.

The maximum capital cost involved would add only 2.4 percent to the present capital value of public water supply systems. However, the overall costs would be increased by 34 percent. If all these measures were applied to a single public water supply, the per capita annual cost to its users would rise $7.81, or 39 percent.

The disease incidence reduction rates are based on estimated reductions from present waterborne disease levels, and they are based solely on personal experience and studies (Hudson, 1962) and on recollections of published accounts.

Table 21

Characteristics of Existing Public Water Supply Systems and of Measures To Improve Water Quality

Item	Costs, Million Dollars		Estimated Annual Benefits, Million Dollars	Estimated Resulting Reduction in Exposure to Waterborne Pathogens, Percent		Estimated Reduction in Concentration of Material, Percent		Annual Per Capita Costs for Those Served
	Capital	Annual		Bacillary	Viral	Exotic Organics	Heavy Metals	
Characteristics of Existing Systems								
Community Water Supplies	70,000	2,500	--	--	--	--	--	$20
Other Public Water Supplies	15,000	1,000	--	--	--	--	--	20
Total	85,000	3,500						
Characteristics of Additional Measures to Improve Water Quality								(Incremental Annual Cost)
Up-grading Salaries	--	660	--					3.34
Training	--	110	--					0.56
Surveillance				90	30	0	0	
State	--	146	--					0.81
Federal	--	25	--					0.14
Limited Chlorination	25	2.25	--					0.75
Universal Chlorination	225	22.5	--	95	40	10	0	0.75
Clarification or Softening	1,650	165	--	100	98	90	90	0.92
Activated Carbon Treatment	36	12	50	100	100	99	99	0.24
Corrosion Control	40	20	210	--	--	--	--	0.30
								1.81

SECTION VII
BENEFIT/COST ANALYSIS OF A DRINKING WATER IMPROVEMENT PROGRAM

The purpose of this chapter is to tie together the findings of the previous sections into a benefit/cost analysis. The specific program to be analyzed here is the four-point program discussed in Section V. This program seems reasonable, embodies many concerns of current legislation, and its components (adequate disinfection, surveillance, presence of a trained operator, presence of a residual chlorine) are all required in order to maintain good water quality.

Costs of the Water Quality Program

Implementation of the program discussed in Section V would incur costs for surveillance, chlorination, and operators. Section VI developed some of these costs. These costs of implementing a water quality program includes Federal and State costs in the categories of fixed (capital) costs and variable (operating) costs.

The EPA (1973, 1973a) has estimated costs of surveillance to be $40 million for the estimated 40,000 community or municipal systems and $10 million for the estimated 30,000 high risk small public systems identified by the EPA. Since current expenditures for surveillance total $10 million, an additional expenditure by states of $30 million (or $40 million, if small high risk systems are covered) for surveillance is required. In addition, federal spending of $21 million in excess of current $4 million expenditures is proposed under HR 10955 to accomplish the surveillance effort.

Additional costs to correct disinfection deficiencies may be estimated using the characteristics of the CWSS systems by size, and the size distribution reported in Hudson and Rodriquez (1970) to extrapolate to the 40,000 community water systems. Table 1 shows data from the CWSS. Table 2 shows the distribution of the 19007 water systems by size class in 1963. To

Table 22. CWSS Sample Characteristics

Size Class	Population Served in Thousands	Distribution percent	% not disinfecting	% inadequate disinfection
1	under 5	78	60	15
2	5-10	8	24	28
3	10-25	6	34	24
4	25-50	4	25	6
5	50-100	1.7	13	13
6	over 100	2.3	0	9

Source: CWSS (1970)

Table 23. Distribution of 19,007 Municipal Water Facilities by Size of Community Served, 1963

Size Class	Population Served	# of System	% of Total Systems
1	under 5,000	15,238	8.0
2	5,000-10,000	1,503	7.9
3	10,000-25,000	1,225	6.4
4	25,000-50,000	573	3.0
5	50,000-100,000	258	1.35
6	over 100,000	210	1.10

Source: Hudson and Rodriguez (1970)

extropolate disinfection characteristics to the 40,000 systems, it is assumed that the new systems coming into existence since 1963 fall in either size class 1 or 2 with 91% in class 1 and 9% in class 2 (this is the same breakdown between class 1 and 2 as in 1963). There, the same characteristics by size class (whether adequate or inadequate disinfection) as in the CWSS we assumed to hold for the 40,000 systems. The results are shown in Table 24.

Table 24. Estimated System Characteristics for 40,000 Community and Municipal Water Supplies

Size Class	Population Served	# of Systems	# not desinfecting	# inadequate disinfection
1	under 5,000	34,341	20,605	515
2	5,000-10,000	3,393	814	950
3	10,000-25,000	1,225	416	294
4	25,000-50,000	573	143	34
5	50,000-100,000	258	33	33
6	over 100,000	210	0	19

Using a capital cost from the previous chapter of $5 per person for disinfection and an assumed cost of $2 per person for upgrading inadequate systems yields Table 25 derived from Table 24. Table 25 shows that correcting deficiencies in disinfection for the 40,000 systems requires about $204 million in capital expenditure. If small high risk systems are included in disinfection requirements at a cost of $1000 per system, an additional cost of $30 million would be incurred. It can be assumed that disinfection facilities would have a lifetime of at least twenty years.

Table 25. Capital Costs for Adding Disinfection and Upgrading Deficiencies in Disinfection 40,000 Systems in the U.S.

Pop. Size Class	Average Pop. Served in Thousands	Cost of Adding Disinfection to Systems not Disinfecting	Cost of Upgrading in Systems With Inadequate Disinfection	Total Capital Cost of Program
1	1	103,925,000	10,302,000	113,237,000
2	5	20,350,000	9,500,000	29,850,000
3	10	20,800,000	5,800,000	26,680,000
4	25	17,875,000	1,700,000	19,575,000
5	50	8,250,000	3,300,000	11,500,000
6	100	0	3,800,000	3,800,000
				204,732,000

Variable costs of disinfection would include cost of chemicals and operators. Additional costs for chemicals can be assumed to be minimal compared to capital costs. Furthermore, it can be assumed that there would be little cost difference whether the residual chlorine requirement was 0.3 or 0.5 ppm residual chlorine. (In computing benefits a residual chlorine of 0.5 ppm will be assumed). Thus, there is no cost included here in the chemicals category.

It is very difficult to estimate the additional costs due to operators required by the program since so little is known about current operator characteristics. (The CWSS does not give consistent information in this area.) An increase in the number of operators may not be necessary if small systems share operators. The need, as pointed out in the previous chapter, is to upgrade operator quality and concomitantly operator salaries. Since adequate data for estimating increased operator salary costs is lacking, these costs will not be included in the benefit/cost analysis here.

The CWSS also showed that most systems lack clarification. Among the size classes, the amount of deficiency ranges from 93% in class 1 to 40% in class 7. The costs of correcting clarification deficiences would then be much greater than those indicated in the previous section. Clarification is no doubt necessary for good water treatment, but it will also not be included in the analysis here due to lack of better information.

Table 26 summarizes costs to be used in the benefit/cost analysis. These include only costs due to disinfection and surveillance since information on costs is lacking in other areas. However, as Section V observed, these two improvements with a residual chlorine of 0.50 ppm alone will reduce SPCD by 100%.

Table 26. Summary of Additional Costs for the Improvement Program

	State		Federal
	Capital	Annual	Annual
1. Disinfection	$704 (234)*		
2. Surveillance		$30 ($40)*	$21

*Including small high risk systems

Benefits of the Water Quality Program

Chapter II developed social costs caused by waterborne salmonellosis, shigellosis, and hepatitis. On a per incident basis, the average cost for each of the three diseases are:

Salmonellosis	$184
Shigellosis	$172
Hepatitis	$1512

These costs include both the cost borne by the individual for treatment, lost wages, etc. and the cost to society due to lost productivity. While the sum of these amounts may involve some double counting in terms of income accounting, the sum represents what employers and workers would be willing to pay to avoid an incident of illness. If a water quality program reduces these three diseases, then each cause of disease eliminated may be valued by the above; that is, the social benefit due to a program is the number of cases eliminated times the social cost above per incident. Of course, there are other diseases (gastroenteritis, amebiasis, etc.) which would also be affected by water quality improvements; these cannot be included here because of lack of information. It must also be pointed out that the above costs do not include doctors' fees or medicine also due to lack of information; however, it is expected that these costs are minimal.

Using 1972 reported disease figures for the U.S. given by the CDC (1973) and the percent disease reduction figures from Section V for 100% reduction in SPCD gives the following reported disease reduction estimates for 1972:

Salmonellosis	33,151 x .249 = 5516
Shigellosis	20,205 x 1 = 20,205
Hepatitis	54,074 x .244 = 13,194

These figures for reported diseases must be inflated by a multiplier to obtain actual disease reductions. Several multipliers were reported in Section IV. It is difficult to say which is the correct multiplier to use in this case since we have already corrected for the disease which is reported but not identified as being due to water. A multiplier of 12.1 attributed to Barker et al. (1969) in Section IV may be used for salmonellosis

and shigellosis. A multiplier of 5 for hepatitis may be calculated as follows from Morse et al. (1972):

$$ND/D = \frac{.5\ (.41)}{.54 + .5\ (.41)} = .26$$

$$D = 4R$$

$$T = D + ND = 4R + .26(4R) = 5R$$

where T = total disease; R = reported disease; D = # going to doctor; ND = # not going to doctor.

These multipliers yield reduction in actual disease of:

Salmonellosis	66,743
Shigellosis	244,480
Hepatitis	65,970

Applying the cost per disease estimates from Chapter II then gives the following benefits for the water quality program:

Salmonellosis		$12,280,712
Shigellosis		42,050,560
Hepatitis		94,746,640
	TOTAL	$154,077,912

This amount will be used as the amount of annual benefits due to the water quality program in the benefit/cost analysis below. (assuming that each reported case means one visit to a doctor at $20 a visit yields and additional social cost of only $778,300 to this amount.)

Benefit/Cost Analysis of the Water Quality Program

Once the annual benefits (B) and the corresponding initial capital costs (K), annual operating costs (C), and the annual federal costs (F) of the program are determined, a cost benefit ratio may be computed as follows (assuming the capital equipment to have a life of twenty years):

$$\text{Ratio} = \frac{B\left(\sum_{t=1}^{20}\right)\frac{1}{(1+i)^t}}{\frac{K}{1+i} + (C+F)\left(\sum_{t=1}^{20}\right)\frac{1}{(1+i)^t}} = \frac{B}{K\left/\left(\sum_{t=0}^{19}\frac{1}{(1+i)^t}\right)\right. + C + F}$$

when "i" denotes the market rate of interest. This formula compares the "present value" of the twenty years of benefits with the "present value" of the twenty years of costs; capital costs occur only in the first year while benefits and annual costs occur every year. The concept of present value is

based on the opportunity cost of capital. If money wasn't used to build a water treatment plant, it could have been used for some other purpose and at the very least invested to receive a return at interest rate i.

Table 27 gives values of the above ratio using an interest rate of 6%. The two values of the ratio depend on whether or not the 30,000 high risk systems are included. (Recall that salary adjustments or additions are not included in these ratios.)

It must be remembered that the benefits in Table 27 represent benefits for the three diseases only. The benefits for the three diseases alone amply cover costs of surveillance and disinfection. Annual Salary adjustment or additions of up to 5% would still make the program have benefits exceed costs in terms of the three diseases. Inclusion of other water-related diseases would give an even better benefit/cost ratio. When the small high risk systems are included, costs increase and the benefit/cost ratio decreases. Including costs for the high risk systems means that some

Table 27. Benefit/Cost Ratios*

1. 40,500 Community Systems	$\frac{154}{\frac{204 + 30 + 21}{12.16}}$	= 2.27
2. Community and High Risk Systems	$\frac{163}{\frac{234 + 40 + 21}{12.16}}$	= 2.03

*not including salary adjustment in costs and including only Salmonellosis, Shigellosis and Hepatitis in benefits.

benefits for these systems should also be added. It is difficult from this study to say by how much benefits would increase since information is available based solely on disease reported by residents covered by community systems. Many of the high risk systems (such as campgrounds) serve non-resident populations. Using an estimate of an average $1.50 benefit per person consistent with the results for the community systems and an estimate of 200 persons served by each of the 30,000 high risk systems yeilds an estimate of $9 million additional benefits for high risk systems for a total benefit of $163 million. The gain in benefits by including these systems is $9 million while the incremental gain in costs is $13 million; this indicates inclusion of these systems is probably not worthwhile using the estimates and assumptions used here. Since the feasibility of including these systems depends heavily on the assumptions about their benefits, it seems this is an area which requires further research.

Finally, it must be noted that costs of drinking water improvement may not be as high as those indicated here. This is because there are economies of scale associated with treatment and experienced operators may be shared by small systems. The extent of savings under consolidation and coordination

of small systems is another area which needs to be explored in future research. Since the same benefits could then be attained at lower costs, inclusion of these considerations in benefit/cost comparisons would mean higher benefit/cost ratios than those shown in Table 27.

SECTION VIII
ACKNOWLEDGEMENTS

The major contributions to the study made by Raphael Samper and Edward Black, graduate assistants, are acknowledged with sincere thanks. Mr. Samper was involved in the processing of the CWSS data, collecting of health data and development and application of the regression models, which will be the basis of his doctoral dissertation Mr. Black used a regression model to evaluate the cardiovascular disease---hard water system.

Dr. Stephen Zellner, of the Medical Center at the University of Florida contributed greatly to the early stages of the study with his knowledge and experience in the public health and epidemilogical fields.

Dr. A. A. Sokoloski, formerly of the Water Supply Division of EPA, helped in the development of the project goals and approaches to those goals. His assistance is greatly appreciated.

The assistance of Dr. Edgar A. Jeffrey, Chief of the Operations & Surveillance Section of the Water Supply Division of the Environmental Protection Agency, throughout the latter stages of the study is acknowledged with sincere thanks.

SECTION IX
REFERENCES

A National Program to Conquer Heart Disease, Cancer and Stroke (1965). Report to the President. The President Commission on Heart Disease, Cancer and Stroke, Volume II.

Anderson, T. W., W. H. LeRiche and J. S. MacKay (1969). Sudden Death and Ischemic Heart Disease Correlation with Hardness of Local Water Supply. New Eng. J. Med., 280:805-807.

Anderson, T. W. and W. H. LeRiche (1971). Sudden Death from Ischemic Heart Disease in Ontario and Its Correlation with Water Hardness and Other Factors. CMA Journal, 105:155-160.

Anonymous (Sept. 16, 1974). Unpublished Communication from Eric A. Johnson.

Askeroff, B., S. A. Schoeder, and P. S. Brachman (1970). Salmonellosis in the United States--A Five-Year Review. American Journal of Epidemiology, 92:1.

Barker, W. H. (1969). Food Borne Disease Surveillance. Unpublished Paper.

Barker, W. H., Jr., J. C. Sagerser, C. V. Hall and B. J. Francis (Nov. 14, 1969). Foodborne Disease Surveillance, Washington State, 1969. Paper presented at the annual meeting of the American Public Health Association, Atlantic City.

Barr, A. J. and J. H. Goodnight (1972). Statistical Analysis System. North Carolina State University.

Bierenbaum, M. L., A. I. Fleischman, J. P. Dunn, H. D. Belk and B. M. Storter (1969). Serum Lipids in Hard and Soft Water Communities. Israel J. Med. Sci., 5:657-660.

Biörck, G., H. Boström and A. Widström (1965). On the Relationship Between Water Hardness and Death Rate in Cardiovascular Diseases. Acta Med. Scan., 178L239-252.

Boström, H. and P. O. Wester (1967). Trace Elements in Drinking Water and Death Rate in Cardiovascular Disease. Acta Med. Scan., 181:465-473.

Barrett-Connor, E. and J. D. Connor (1968). Shigellosis Recent Experience with 330 Cases. J. Florida Med. Assoc., 55:540-544.

Bauer, H. (1973). Growing Problem of Salmonellosis in Modern Society. Med., 52:322-330.

Census of Population, 1960. U. S. Bureau of Census.

Center for Disease Control (1971). Morbidity and Mortality. U.S. Department of Health, Education and Welfare. Weekly Report, Annual Supplement Summary, 20:53.

Center for Disease Control (1971). Shigella Surveillance Report No. 27. U. S. Public Health Service. Center for Disease Control, Atlanta.

Center for Disease Control (1973a). Hepatitis Surveillance, Report No. 36. U. S. Public Health Service. Center for Disease Control, Atlanta.

Center for Disease Control (1973b). Foodborne Outbreaks, Annual Summary 1972. U. S. Public Health Service. Center for Disease Control, Atlanta.

Cherubin, C. E., T. Fodor, L. I. Denmark, C. S. Master, H. T. Fuerst and J. W. Winter (1969). Symptoms, Septicemia and Death in Salmonellosis. Amer. J. Epidemiol., 90:285-291.

Commission on Professional and Hospital Activities (1973). Hospital Record Study. Commission on Professional and Hospital Activities, Ann Arbor.

Committee on Salmonella (1969). An Evaluation of the Salmonella Problem. National Academy of Sciences, Washington, D. C.

Comstock, G. W. (1971). Fatal Arteriosclerotic Heart Disease, Water Hardness at Home and Socioeconomic Characteristics. Am. J. Epidem., 94:1.

Cookson, J. T. and North, W. (Jan. 1967; Feb. 1967). Adsorption of Viruses on Activated Carbon. Env. Sci. and Tech., 1:46 and 1:157, respectively.

Correa, P., J. P. Strong (1972). Atherosclerosis and the Geochemical Environment: A Critical Review. Ann. N. Y. Acad. Sci., 199:217-228.

Craun, G. F. and L. J. McCabe (1971). Waterborne Disease Outbreaks, 1961-1970. Presented at the Annual Meeting of the American Water Works Association.

Craun, G. F. and L. J. McCabe (1972a). Review of the Causes of Waterborne-Disease Outbreaks. J. Amer. Water Works Assoc., 75:74-84.

Craun, G. F. and L. J. McCabe (1973b). Summary of Waterborne-Disease Outbreaks in the United States During 1971 and 1972. Presented at the Annual Meeting of the American Water Works Association.

Crawford, M. D., M. J. Gardner and J. N. Morris (1968). Mortality and Hardness of Local Water-Supplies. Lancet., i:827-831.

Crawford, T. and M. D. Crawford. Prevalence and Pathological Changes of Ischemic Heart-Disease in a Hard-Water and in a Soft-Water Area.

Crawford, M. D. and T. Crawford (1969). Lead Content of Bones in a Soft and a Hard Water Area. Lancet., i:699-701.

Crawford, M. D., M. J. Gardner and J. N. Morris (1971). Changes in Water Hardness and Local Death Rates. Lancet., ii:327-329.

Crawford, M. D., M. J. Gardner and J. N. Morris (1971). Cardiovascular Diseases and the Mineral Content of Drinking Water. Br. Med. Bull., 27:21-24.

Dudley, E. F., R. A. Beldin and B. C. Johnson (1969). Climate, Water Hardness and Coronary Heart Disease. J. Chron. Dis., 22:25-48.

Duffey, E. A. and R. E. Carroll (1967). United States Metropolitan Mortality, 1959-1961. National Center for Air Pollution Control, Cincinnati. PHS Publication No. 999 Ap 39.

Durfor, C. N. and E. Becker (1964). Public Water Supplies of the 100 Largest Cities in the United States. 1962 Geological Survey Water-Supply Paper 1812. U.S. Government Printing Office, Washington, D. C.

Dyer, C. H. (July 1974). Manpower: The Important Element in Providing Quality Water Service. Jour. AWWA, 66:7, p 434.

Goldfield, E. (major investigator). County City Data Book, 1962. Bureau of Census, U.S. Department of Commerce.

Eichoff, T. C., P. S. Brachman, J. V. Bennett andJ. F. Brown (1969). Surveillance of Nosocomial Infections in Community Hospitals: I. Surveillance Methods, Effectiveness and Initial Results. J. Infect. Dis., 120:305-317.

Enterline, P. E., A. E. Rikli, H. I. Sauer and M. Hyman (1960). Death Rates for Coronary Heart Disease in Metropolitan and Other Areas. Pub. Health Rep. 75:759-766.

Environmental Protection Agency (Jan. 1973). Safe Drinking Water Act of 1973: Estimates of Benefits and Costs. Unpublished report.

Environmental Protection Agency (1973a). A Rationale for Developing an Estimate of State Public Water Supply Program Costs. Unpublished report.

Epstein, F. H. (1965). The Epidemiology of Coronary Heart Disease. J. Chron. Dis., 18:735-774.

Gardner, M. J., D. D. Crawford and J. N. Morris (1969). Patterns of Mortality in Middle and Early Old Age in the County Boroughs of England and Wales. Br. J. Prev. Soc. Med., 23:133.

Garibaldi, J. A., F. B. Murphy, III and B. T. Wood (1972). Infectious Hepatitis Outbreak Associated with Oak Water. CDC Health Rep., 87:365-370.

Goldberger, J. S. (1966). Economic Theory. Wiley and Sons, New York.

Groover, M. B. and J. C. Culpepper. Death Rates Following A Sudden Change in Hardness of Drinking Water.

Hamer, R. (Dec. 1973). Granular Carbon Filters for Taste and Odor Removal. J. Amer. Water, 65:1, p 176.

Hudson, H. E., Jr. (Oct. 1962). High-quality Production and Viral Removal. Jour. AWWA, 54:10, p 1265.

Hudson, H. E., Jr. and Rodrigues, F. (1970). Water Utility Personnel Statistics. Jour. AWWA, 62:8, p 485 (August).

Hudson, H. E. (Apr. 1966). Loss of Water Main Capacity. Jour. SE Sect. AWWA, 30:1, p 44.

Frosta, J. (1971). Elements of Econometrics. The Macmillan Company, New York.

Kobyashi, J. (1957). On Geographical Relationship Between the Chemical Nature of River Water and Death Rate from Apoplexy. Ber. Ohara Inst. Landw. Forsch., 11:12-21.

Krugman, S., R. Ward and J. F. Giles (1962). The Natural History of Infectious Hepatitis. Amer. J. Med., 32:717-728.

Lamm, S. H., A. Taylor, Jr., E. J. Cangarosa, R. M. Anderson, W. Young, M. H. Clark and A. A. Bruce (1972). Turtle-Associated Salmonellosis: I. An Estimation of the Magnitude of the Problem in the United States, 1970-1971. Amer. J. Epidemiol., 95:511-528.

Lehr, I., H. E. Eisenmer and S. H. Rosenman (1973). A Sociobiological Approach to the Study of Coronary Heart Disease. J. Chron. Dis. 26:13-30.

Lindeman, R. D. and J. E. Isaacson (1964). Correlations Between Water Hardness and Cardiovascular Deaths in Oklahoma Counties. Am. J. Pub. Health. 54:1072-1077.

Kobel, H. O., A. L. Bisno, M. Goldfield and J. E. Prier (1969). A Waterborne Epidemic of Gastroenteritis with Secondary Person-to-Person Spread. Amer. J. Epidemiol., 89:384-392.

Masironi, R. (1969). Trace Elements and Cardiovascular Diseases. Bull. WHO, 40:305-312.

Masironi, R. (1970). Cardiovascular Mortality in Relation to Radioactivity and Hardness of Local Water Supplies in the USA. Bull. WHO, 43.687-697.

Masironi, R., A. T. Miesch, M. D. Crawford and E. I. Hamilton (1973). Geochemical Environments, Trace Elements and Cardiovascular Diseases. Pan. Am. Hlth. Org. Bull., 1:53-64.

McCall, R. (June 13, 1974). Address. American Academy of Environmental Engineers Boston, Mass.

Monti, R. and Silberman, P. (May 1974). Wastewater System Alternates: What are They . . . and What Cost? Water and Wastes Engineering, 11: 5, p 40.

Moriyama, I. M., D. F. Krueger and J. Stamler (1971). Cardiovascular Diseases in the United States. Harvard University Press, Cambridge.

Morris, J. N., M. D. Crawford and J. A. Heady (Apr. 22, 1961). Hardness of Local Water Supplies and Mortality from Cardiovascular Disease. Lancet., 860-862.

Morse, L. J., J. A. Bryan, J. P. Hurley, J. F. Murphy, T. F. O'Brien and W. E. C. Wacker (1972b). An Epidemiologic Report of the Holy Cross College Football Team Hepatitis Outbreak. J. Environ. Health, 34.

Morse, L. J. et al. (1972). The Holy Cross College Football Team Hepatitis Outbreak. The American Medical Association, 219.

Mulcahy, R. (1964). The Influence of Water Hardness and Rainfall on the Incidence of Cardiovascular and Cerebrovascular Mortality in Ireland. J. Irish Med. Assoc., 55:17-18.

Mulcahy, R. (1966). The Influence of Water Hardness and Rainfall on Cardiovascular and Cerebrovascular Mortality in Ireland. J. Irish Med. Assoc., 59:14.

Muss, D. L. (Nov. 1962). Relationship between Water Quality and Deaths from Cardiovascular Disease. JAWWA, 1371-1378.

National Institutes of Health (June 1971). Atheriosclerosis: A Report by the National Heart and Lung Institute Task Force on Atherosclerosis, Vol. II. DHEW, Pub. No. (NIH) 72-219.

Neri, L. C., D. Hewitt and J. S. Mandel (1971). Risk of Sudden Death in Soft Water Areas. J. Epidem., 94:101-104.

Neri, L. C., J. S. Mandell and D. Hewitt (1972). Relation between Mortality and Water Hardness in Canada. Lancet. 1:931-934.

Perry, H. M., Jr. (1972). Hypertension and the Geochemical Environment. Ann. N. Y. Acad. Sci., 199:202-216.

Peterson, D. R., D. J. Thompson and Jun-Mo Nam (1979). Water Hardness, Arteriosclerotic Heart Disease and Sudden Death. Am. J. Epidem., 92: 90-93.

Prest, A. R. and R. Turvey (1966). Costs-Benefit Analysis: A Survey. Surveys of Economic Theory, III. St. Martin's Press, New York.

Pyatt, E. E. and P. P. Rogers (1962). On Estimating Benefit-Cost Ratios for Water Supply Investments. Am. J. Pub. Health, 52:1729-1742.

Reller, L. B., E. J. Gangarosa and P. S. Brachman (1970). Shigellosis in the United States: Five-Year Review of Nationwide Surveillance, 1964-1968. Amer. J. Epidemiol., 91:161-169.

Repsher, L. H. and R. K. Freebern (1969). Effects of Early and Vigorous Exercise on Recovery from Infectious Hepatitis. N. Engl. J. Med., 281:1393-1396.

Rice, D. P. (1966). Estimating the Cost of Illness. Health Economics Series No. 6, U.S. Public Health Service, Washington, D. C.

Robertson, J. S. (1969). The Water Story. Lancet., 6:1160.

Sauer, H. I. and P. E. Enterline (1959). Are Geographic Variations in Death Rates for Cardiovascular Diseases Real. J. Chron. Dis., 10:513-524.

Sauer, H. I. (1962). Epidemiology of Cardiovascular Mortality--Geographic and Ethnic. AJPH, 52:94-105.

Sauer, H. I. (May 19 to June 26, 1970). Geographic Differences in the Risk of Premature Death. Business and Government Review.

Sauer, H. I., D. W. Parke and M. L. Neill (1973). Associations between Drinking Water and Death Rates. Pam. Am. Health Org. Bull., 7:318-325.

Sauer, H. I. (1974). Relationship between Trace Element Content of the Drinking Water and Chronic Diseases. Presented at the 16th Water Quality Conference, University of Illinois at Urbana, Ill.

Schroeder, H. A. (1958). Degenerative Cardiovascular Disease in the Orient. J. Chron. Dis., 8:312-333.

Schroeder, H. A. (1960). Relation between Mortality from Cardiovascular Disease and Treated Water Supplies. JAMA, 172:1902-1908.

Schroeder, H. A. (1966). Municipal Drinking Water and Cardiovascular Death Rates. JAMA, 195:81-85.

Schroeder, H. A. and L. A. Kraemer (1974). Cardiovascular Mortality, Municipal Water and Corrosion. Arch. of Environ. Health, 28:303-311.

Sigworth, E. A. and Smith, S. B. (June 1972). Adsorption of Inorganic Compounds by Activated Carbon. Jour. AWWA, 64:386

Stamler, J. (1962). Cardiovascular Diseases in the United States. Am. J. Card., 10:319-340.

Stitt, F. W., D. G. Clayton, M. D. Crawford and J. N. Morris (1973). Clinical and Biochemical Indicators of Cardiovascular Disease Among Men Living in Hard and Soft Water Areas. Lancet., i:122.

Strong, J. P., P. Correa and L. A. Solberg (1968). Water Hardness and Atherosclerosis. Lab. Invest., 18:620-622.

Thursby, R., K. McClelland and B. Poole (May 1974). Staffing and Budgetary Guidelines for State Drinking Water Supply Agencies. National San. Found., Ann Arbor, Mich.

U. S. Department of Commerce, Bureau of the Census (1970). Compendium of Government Finances, 1967 Census of Government, Vol. 4, No. 5, Washington, D. C.

U. S. Department of Commerce, Bureau of Census, Social and Economic Statistical Administration (1973). County and City Date Book 1972. A Statistical Abstract Supplement, Washington, D. C.

U. S. Department of Commerce (1973). Survey of Current Business. 53:15 (May).

U. S. Department of Health, Education and Welfare, Public Health Service (July 1970). Community Water Supply Study, Analysis of National Supply Findings.

U. S. Public Health Service (1973). Morbidity and Mortality Weekly Report. Annual Supplement, 21:53.

U. S. Weather Bureau (1959-1961). Climatological Data. U.S. Department of Commerce, Washington, 46-48.

Wacker, W. E. C., J. F. Blodan, P. J. Snodgrass, L. W. Chang, L. J. Horne, T. F. O'Brien and W. J. Reddy (1972). The Holy Cross Hepatitis Outbreak, Arch. Intern. Med. 130:357-360.

World Health Organization (1955). Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death. World Health Organization, Geneva, 1957.

SECTION X
PUBLICATIONS

One paper has been submitted for publication.

Measuring the Benefits of Drinking Water Quality Improvements, Edna T. Loehman and Raphael Sanger, submitted to American Agricultural Economics Association.

APPENDIX A
DATA DESCRIPTION

The following will describe the variables used in each of the two types of regressions (disease and water quality).

Disease Equations

A. Disease Rates

Reported cases of infectious hepatitis, shigellosis and salmonellosis were collected by phone from state and/or county health departments by counties for the 51 counties covered by the CWSS [8] and the city of N.Y. The disease data is for 1969, the same year in which the CWSS data was collected. (It had been hoped that it would be possible to collect disease data by census tract and in this way have a close correspondence of disease to water system. However, disease data is only reported by county totals.) The actual number of cases reported per county for each of these diseases were divided by county populations for 1969 to arrive at rates of disease per 1000 population for HEP,RSAL and RSAL respectively.

Some adjustments were made in this disease data: for infectious hepatitis it was found that counties in Kentucky and Louisiana included both infectious and serum hepatitis in their reports. For the three counties in Kentucky it was not possible to determine the infectious hepatitis separately so data from these counties was not used. In the case of Louisiana the average ratio of infectious to total hepatitis by county for the period 1970-1972 was used (with Dr. Zellner's approval) to estimate infectious hepatitis from total hepatitis for the year 1969. Also, for Orange County, Vermont, it was found that 26 out of 27 cases of shigellosis reported were accounted for by an outbreak at a private camp which was not under any community water supply; the disease reported was adjusted in this case. Table A1 shows the rates of reported cases per thousand population of infectious hepatitis, shigellosis and salmonellosis, for the 58 counties and the city of New York used in the regression. (In the case of shigellosis and salmonellosis, zero reported cases is converted to .0001 because of later conversion to logarithms.)

B. Water Quality

The water quality data used (Table A1) was obtained from the CWSS. Water quality was measured as a standard plate count of bacterial per ml. of water x 100. Each water system in the CWSS had water quality observations for both water in the distribution system (SPCD) and raw water (SPCR). The SPCD (or SPCR) for a county is the average of SPCD (or SPCR) values of all systems in that county. (Counties that had water quality samples with bacterial too numerous to count were assigned a value equal to the highest value found in the systems in the county.)

C. Socio-Economic Variables

The data on population per square mile (PPM), per capita income (PPY), and percent urban population (URB) for 1970 were taken directly from the 1972 County and City Data Book of the Bureau of the Census [7]. The data on percent population between the ages of five and twenty-five (PFTF) and percent population five and below, found in the County and City Data Book [7]. The percent population served (PPS) was computed using the total population served by systems in a county as reported by the CWSS and divided by the county population (1970) as found in the County and City Data Book [7]. Per capita state and local health expenditure for each county (SLPC) was computed from data in the 1967 Census of Government [6], and is the sum of state health expenditure per capita plus the county per capita health expenditure for 1967.

D. Time Variables

The water quality data from the CWSS includes dates when samples were taken. All samples were collected between the third and nineth months of the calendar year. Counties were classified according to three time groups: T1 denotes samples taken predominatly in the third, fourth and fifth months, T2 denotes samples taken predominantly in the sixth month and T3 denotes samples collected predominantly in the seventh, eight and nineth months. These time variables are assigned zero-one values to indicate in which period water quality tests occurred. The purpose of these variables is to account for possible differences in measurement techniques and temperature factors among the three time periods.

Water Quality Equation

The CWSS study of 1969 [8] was selected as the source of water quality and treatment characteristics data since it provided the most complete information and the largest sample available. This study examined 969 water supply systems in eight SMSA's and the state of Vermont, covering 42 counties in total. The data was made available by the Water Division of EPA on two magnetic tapes: Tape 1 contains the results of 4167 grab samples of water taken throughout the 969 systems (2595 samples were taken in distribution systems, the remainder at the source or treatment plant); Tape 2 contains the responses to a questionaire on treatment plant and operator characteristics, taken for each of the systems.

A. Water Quality

The most complete data on water quality (Tape 1) consisted of standard plate counts per milliliter of water. Using an eleven digit indentification number identifying each sample within a system, it was possible to determine where samples were taken (i.e. treatment plant, source or distribution system) and whether the samples were of raw water or finished water. 98 systems (these systems were not excluded in the disease equation because only distribution information was required) were excluded because they did not have both raw water and distribution system water

samples; one system had no water samples. For each of the remaining 670 systems, the respective raw water samples and distribution system samples were averaged to provide an indicator of raw water quality (SPCR) and distribution system water (SPCD). There were 56 systems which had a ground water source, provided no treatment and had samples taken in the distribution system only; in these systems SPCR was set equal to SPCD.

In averaging SPCD and SPCR for each system it was found that some had observations reported as 99999 indicating bacterial counts too numerous to count. Instead of deleting such systems from computations of SPCR and SPCD, this value was replaced by the value of the sample with the highest bacterial count in the respective system. To permit logarithmic computations, it was also necessary to transform all values of zero to .0001; the water quality variables in log form are LSPCD and LSPCR for distribution and raw water quality respectively.

B. Water Source

The eleven digit sample identification number on tape 2 also identifies the source of the raw water according to surface, ground or mixed. Three 0-1 variables (SURF, GRND, MIX respectively) are used to indicate the source of raw water. Thus for example, if the source of water of a system were surface, SURF has a value of 1, and GRND and MIX both have values of zero.

C. Surveillance

Tape 2 contains the answers to several questions used to determine the compliance with the bacteriological surveillance program over the past 12 months. The surveillance index used (SURV) consists of the sum of the number of months when the quality of water in the distribution system did not meet the standards, plus the number of months when less than the required number of samples were taken, plus the number of months when no samples were collected. The index which resulted ranged from 0 to 36, from complete compliance to no compliance at all. In order to facilitate interpretation and permit logarithmic conversion, this index was then changed to range from 1 to 37, and the larger the number the greater the degree of compliance (37 represents full compliance). The log of SURV is denoted as LSURV.

D. Treatment

Tape 2 contains the information on treatment. The questionaire was directed at determining the adequacy of the treatment process; answers of yes or no indicated whether a process was adequate or not. Where no response to these questions was indicated on the tape, it was assumed that the process was not present (this assumption is not always correct as the data on the tape shows). This provided three alternative situations in terms of the treatment process: the process was present and adequate, present and inadequate, or not present. The first alternative was considered the most preferable and was assigned a value of 3, the second alternative was assigned 2, and the least preferable alternative, that of a process not being present, was assigned the number 1.

Of the treatment processes reviewed by the questionaire, five were considered to be important in terms of determining the bacteriological quality of water (flocculation [FLOC], settling [SED], filtration [FILT], disinfection [DIS], capacity and booster chlorination [CLOR]). Those five variables were each assigned a 1, 2, or 3 depending on the adequacy and presence of the process.

Using the above data, flocculation, settling and filtration were combined into a single variable called clarification (CLAR). To define the variable CLAR, the maximum value of FLOC, SED, FILT was taken; thus if any one of these processes was adequate, clarification was indicated as being adequate.

One problem was that many systems did not report clarification processes. The engineers pointed out that ground water may not always require clarification so that assuming CLAR = 1 in such cases may be over-severe. For systems with a ground water source and no clarification information, a CLAR value of 2 (process present but inadequate) was assigned.

The disinfection variable (DIS) also has values from 1 to 3 in ascending order as the process goes from absent, to present and adequate. The booster chlorination variable (CLOR) was not used in the final regression because of the large number of blanks in the reported data.

E. Education

The questionaire used to collect the information on Tape 2 asked the level of formal education received by the plant operator and the level of training obtained in sanitary microbiology, water treatment and/or water chemistry. Data on level of training were not on the tape and so could not be used. In terms of highest level of formal education attained by an operator (ED), four choices were available: 8th grade or less, high school, technical or trade school, and university. These alternatives were assigned values of 1, 2, 3 and 4 respectively. Where the presence of an operator was not indicated (see Position) it was found that no answer was provided for education level either; a value of one was assigned to these cases.

F. Position

Tape 2 contained a "yes-no" answer to the questions of whether or not the operator was a full-time employee. Where the answer was "yes," a 3 was assigned to the variable (POS) and a "no" answer was assigned a 2; a blank response meant that at most, an occasional operator was present and a value of 1 was assigned to these cases.

G. Pressure

Tape 1 provided information on pressure as recorded at the time of taking the water quality samples. Observations were available for only 291 of the 870 systems, thus this variable was not used.

H. Residual Chlorine

Tape 1 provides information on the residual chlorine (RC) in the samples collected. All systems did not have a value for this variable; where no value was present a zero was inserted (changed to .001 to facilitate computation). Where more than one sample for a system had a value for residual chlorine, RC is the average residual chlorine count. The variable is measured in ppm times 100.

I. Time

Systems were grouped into three time periods in the same way as in the disease regression. All systems where samples were collected predominantly in the third, fourth and fifth months were assigned under a T1 value of one, systems in counties where samples were collected predominantly in the sixth month had T2 = 1, and those in the seventh, eight and ninth months had T3 = 1.

TABLE A1 COUNTY DISEASE RATES

IC	RREP	RSHI	RSAL
9[illegible]033	0.31[illegible]	0.0[illegible]	0.0[illegible]
3[illegible]036	0.[illegible]	0.[illegible]	0.[illegible]
6[illegible]031	0.[illegible]	0.[illegible]	0.[illegible]
5[illegible]15	0.[illegible]	0.[illegible]	0.[illegible]
7[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
7[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
6[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
6[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
6[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
6[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
7[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
7[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
7[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
7[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
6[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
2[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
6[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
2[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
2[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
5[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
5[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
4[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
3[illegible]	0.[illegible]	0.[illegible]	0.[illegible]
1

TABLE A2 COUNTY WATER QUALITY ECONOMIC AND SOCIOLOGICAL DATA

IC	SPCO	PPM	SLPC	PF TP	PZE	PRY
010732	1016.6459	64.0000	6.8200	0.3020	0.2390	3084.0000
020230	849.6018	34.0000	5.7300	0.3014	0.2960	3072.0000
020701	932.7200	49.0000	5.2300	0.2924	0.2620	2841.0000
010713	3.4950	61.0000	1.6100	0.3459	0.3700	2547.0000
020715	1150.6400	462.0000	3.4200	0.3010	0.3110	4042.0000
020105	111.0150	1229.0000	4.5200	0.3789	0.3000	2834.0000
020730	909.1430	617.0000	4.1910	0.3014	0.2980	3026.0000
020706	84.0000	3013.0000	3.7300	0.3717	0.3320	3770.0000
020744	1207.1000	100.0000	4.7200	0.3172	0.3000	2730.0000
020732	7.3000	72.0000	2.1900	0.3042	0.3120	3095.0000
020715	554.7200	57.0000	2.8600	0.3002	0.2910	2864.0000
020734	173.0830	305.0000	3.3000	0.3344	0.3270	3410.0000
020748	970.7400	1000.0000	4.4100	0.3502	0.3330	3441.0000
020703	53.7900	75.0000	5.0200	0.3743	0.3270	3441.0000
020731	1.4230	2032.0000	11.0200	0.3174	0.3040	3221.0000
020713	27.4251	4504.0000	6.0400	0.2786	0.2440	4044.0000
020744	124.9800	1200.0000	8.1700	0.3717	0.3400	3791.0000
020752	85.0750	1813.0000	5.8400	0.3823	0.3470	3320.0000
020707	83.8530	2219.0000	7.8300	0.3334	0.3160	3501.0000
020713	111.3310	2034.0000	2.6200	0.4037	0.3900	4514.0000
020731	80.8482	2220.0000	4.3100	0.3734	0.3460	3186.0000
020702	151.2990	210.0000	2.0100	0.4147	0.3900	2803.0000
020715	107.7412	81.0000	3.9600	0.4054	0.3810	1742.0000
020710	328.0703	2044.0000	4.7400	0.3447	0.3210	3098.0000
020732	420.0194	258.0000	9.6200	0.3832	0.3300	3400.0000
020711	921.7859	31.0000	6.3800	0.4000	0.3600	3400.0000
020702	284.8298	44.0000	6.7500	0.3514	0.3240	4272.0000
020733	41.2180	27.0000	6.5700	0.3719	0.3410	4041.0000
020734	1121.6999	181.0000	7.0000	0.4359	0.3850	3773.0000
020705	426.8750	8.0000	6.3200	0.3823	0.3460	3276.0000
020706	10.3110	87.0000	6.4200	0.3700	0.3400	3807.0000
020707	30.5200	43.0000	5.2700	0.3544	0.3210	3484.0000
020708	78.7200	28.0000	6.3100	0.3974	0.3500	3400.0000
020709	95.1470	26.0000	6.3800	0.3743	0.3410	3441.0000
020710	369.2530	28.0000	6.6300	0.3764	0.3740	4041.0000
020711	30.3330	57.0000	6.3500	0.3729	0.3410	2171.0000
020712	32.3340	67.0000	6.6700	0.3788	0.3437	3469.0000
020713	107.8410	43.0000	6.2600	0.3614	0.3320	3438.0000
020714	193.6200	40.0000	6.5200	0.3438	0.3400	3838.0000

TABLE A2 COUNTY WATER QUALITY ECONOMIC AND SOCIOLOGICAL DATA

Reproduced from best available copy.

IC	PDS	URE	T1	T2	T3
30[illegible]33	0.0372	0.7800	1.0000	0.0	0.0
9[illegible]320	0.[illegible]80	0.[illegible]60	1.0000	0.0	0.0
3[illegible]91	0.[illegible]91	0.[illegible]70	0.0	0.0	0.0
5[illegible]15	0.[illegible]14	0.[illegible]30	1.0000	1.0000	0.0
7[illegible]00	0.9543	0.[illegible]92	0.0	0.0	0.0
7[illegible]33	0.[illegible]23	0.[illegible]24	0.0	1.0000	0.0
6[illegible]20	1.0250	0.[illegible]30	0.0	1.0000	0.0
6[illegible]30	1.[illegible]21	0.[illegible]75	0.0	1.0000	0.0
6[illegible]34	1.1013	0.[illegible]10	0.0	1.0000	0.0
6[illegible]02	0.[illegible]84	0.[illegible]60	0.0	1.0000	0.0
7[illegible]11	0.7274	0.[illegible]70	0.0	1.0000	0.0
7[illegible]24	0.[illegible]14	0.[illegible]20	0.0	1.0000	0.0
7[illegible]48	1.1234	0.[illegible]70	0.0	1.0000	0.0
7[illegible]32	0.[illegible]61	0.[illegible]30	0.0	1.0000	0.0
2[illegible]21	1.1237	1.[illegible]00	0.0	0.0	1.0000
2[illegible]30	0.[illegible]73	0.[illegible]70	0.0	0.0	1.0000
3[illegible]46	0.[illegible]14	0.[illegible]20	1.0000	0.0	0.0
2[illegible]26	0.7379	0.[illegible]86	0.0	0.0	1.0000
2[illegible]00	1.0240	0.[illegible]30	0.0	1.0000	0.0
2[illegible]12	0.[illegible]73	0.[illegible]60	1.0000	0.0	0.0
2[illegible]41	0.[illegible]33	0.6610	1.0000	0.0	0.0
2[illegible]43	0.[illegible]45	0.[illegible]20	1.0000	0.0	0.0
2[illegible]35	0.[illegible]19	0.[illegible]10	1.0000	0.0	0.0
4[illegible]20	0.[illegible]21	0.[illegible]20	1.0000	0.0	0.0
2[illegible]20	0.6028	0.[illegible]50	1.0000	0.0	0.0
1[illegible]04	0.[illegible]26	0.[illegible]01	0.0	0.0	1.0000
1[illegible]02	0.[illegible]11	0.[illegible]10	0.0	0.0	1.0000
1[illegible]04	0.2562	0.[illegible]01	0.0	0.0	1.0000
1[illegible]04	0.5717	0.[illegible]10	0.0	0.0	1.0000
1[illegible]08	0.[illegible]28	0.[illegible]01	0.0	0.0	1.0000
1[illegible]06	0.[illegible]04	0.[illegible]20	0.0	0.0	1.0000
1[illegible]32	0.[illegible]32	0.[illegible]01	0.0	0.0	1.0000
1[illegible]08	0.4702	0.[illegible]01	0.0	0.0	1.0000
1[illegible]36	0.3475	0.[illegible]01	0.0	0.0	1.0000
1[illegible]10	0.[illegible]01	0.[illegible]10	0.0	0.0	1.0000
1[illegible]11	0.[illegible]25	0.[illegible]70	0.0	0.0	1.0000
1[illegible]12	0.[illegible]01	0.[illegible]30	0.0	0.0	1.0000
1[illegible]13	0.[illegible]70	0.[illegible]30	0.0	0.0	1.0000
1[illegible]14	0.[illegible]60	0.1280	0.0	0.0	1.0000

APPENDIX B
TAPE DESCRIPTION

From the two CWSS tapes mentioned in Appendix A, a single tape was prepared for use in estimating the regressions reported here; this was much more convenient that using two tapes. Since others may wish to use this tape also, the following describes the tape and location of data on it.

BYTES	DESCRIPTION
1 -11	System I.D. numbers
13	Mixed water source = 1, 0 otherwise
14	Ground water source = 1, 0 otherwise
15	Surface water source = 1, 0 otherwise
17-25	Standard plate count for distribution system water (SPCD)
27 - 35	Standard plate count for raw water (SPCR)
37 - 43	Parts of residual chlorine per 100 milliliters
44 - 50	Water pressure in the distribution system
52 - 53	Number of months the Drinking Water Standards were not met during the last 12 months due to quality. (Item 6-A, number 4-a from questionaire, Appendix C.)
55 - 56	Number of months when samples taken were less than those recommended by the Drinking Water Standards. (Item 6-A, number 4-b from questionaire, Appendix C.)
58 - 59	Number of months when no samples were collected. (Item 6-A, number 4-c from questionaire, Appendix C.)
62 - 63	SURV-sum of the number of months standards not met, number of months when samples taken were less than those recommended and number of months no samples were taken. (0 means full compliance, 36 means complete failure to comply.)

65	Flocculation	1 = present and adequate
67	Settling	2 = present and inadequate
69	Filtration	3 = not present
71	Disinfection	(Item 9-A, numbers 4-3, 4-f, 4-h, 4-i, and 5-b.)
73	Booster chlorination	
76 - 79	Rated capacity of system in million gallons per day	
81 - 85	Maximum capacity of system in million gallons per day	
87 - 91	Average capacity of system in million gallons per day	
93 - 95	Total cost of production in cents/1,000 gallons (Item 11-H, number 5)	
97	Highest level of formal education attained by operator 0 = no operator, 1 = eight grade or less, 2 = high school, 3 = technical or trade school, 4 = university. (Item 10-A, number 1)	
99	Position of operator: 1 = full-time, 2 = part-time, 3 = no operator. (Item 10-A, number 8)	
101 - 107	Population served in 1969	
109	Class of system by population of community it serves: 0 = unknown, 1 = <1000, 2 = 1000-5000, 3 = 5000-10,000, 4 = 10,000-25,000, 5 = 25,000-50,000, 6 = 50,000-100,000, 7 = 100,000-250,000, 8 = 250,000-500,000, 9 = >500,000	
111	Cross-connection control program through re-inspection. 1 = yes, 2 = no. (Item 8-C, number 2).	
113	If system owned by a community, i.e. public funds, then equals 1, zero otherwise.	
115	If system is investor owned then equals 1, zero otherwise.	
117	If system is owned by a trailer park, institution or recreational area, then equals 1, zero otherwise.	

120	Operator classification: 5 = engineer, 4 = part-time and not certified, 3 = part-time and certified, 2 = full-time and not certified, 1 = full-time and certified, 0 = no operator. (Source of this was not the CWSS questionaire.)

APPENDIX C
CWSS QUESTIONAIRE

Reproduced from best available copy.

MUNICIPAL WATER SUPPLY SANITARY SURVEY

BOB #85-S69004
Exp. March 1970

1. Leave this blank (for office use only) SURVEY DATE Date of your visit (mo. day yr.)

2. Name of supply Same as on inventory

3. Location Same as on inventory
post office common name, if different

4. Demands (GPD)

	PRESENT	10-YR. ESTIMATE	UNKNOWN
A. Avg. day	Empty		
B. Max. day	Empty		
C. Max. month	Empty		

5. Water use has been restricted ☐☐ times for a total of ☐☐☐ days during any one year of the past 5 years.

6. LABORATORY CONTROL

A. Bacteriological (Distribution system only)

(1) Min. number samples recommended per month by PHS DWS → From Fig. 1- D.W.S. ☐☐☐ UNKNOWN ☐

(2) Avg. number/month for last 12 months May be ☐☐☐ ☐

(3) Range of least and most monthly samples from Empty to ☐☐☐ ☐

(4) Number of months the Drinking Water Standards were not met during the last 12 months for: No. of months END CARD ONE UNKNOWN

(a) Quality — See 5.2- D.W.S. ☐☐ ☐

(b) Number of samples (Months when sample numbers < no. in 6.A.(1) ☐☐ ☐

(c) NONE collected No samples for ___ months. ☐☐ ☐

(5) Are samples representative of distribution system? Judge from map of distribution system. ☐ yes ☐ no

(6) Are check samples collected as provided for in the Drinking Water Standards? See 5.15- D.W.S. ☐ yes ☐ no

(7) Are samples requiring check samples reported by telephone? ☐ yes ☐ no

(8) Is the laboratory certified? ☐ yes ☐ no

(a) Within the past 3 years? From operator ☐ yes ☐ no

(b) If answer to one or both, by whom was it certified. ☐ State ☐ PHS

(9) Are samples received by lab within 30 hours? ☐ yes ☐ no

From plant records

Name on each page

B. Chemical (finished water only) 2

(1) Samples of finished water are analyzed each ☐ month, ☐ year, ☐ 2 years, ☐ 3 years, ☐ infrequently ☐ never.

(2) Type of analysis: ☐ complete (CFS) ☐ partial, If any consituent missed

(3) Date of last chemical analysis ☐☐ mo. ☐☐ day ☐☐ yr.

(4) Analyzed by ☐ utility, ☐ state, ☐ PHS, ☐ university, ☐ other.

(5) Tests run for operational control and their frequency are:

Tests	Frequency				
	Continuous	Each shift	Daily	Weekly	Less frequently than weekly
Alkalinity	☐	☐	☐	☐	☐
Aluminum	☐	☐	☐	☐	☐
Chloride	☐	☐	☐	☐	☐
Chlorine residual	☐	☐	☐	☐	☐
Color	☐	☐	☐	☐	☐
Fluoride	☐	☐	☐	☐	☐
Hardness	☐	☐	☐	☐	☐
Iron	☐	☐	☐	☐	☐
Jar tests	☐	☐	☐	☐	☐
Manganese	☐	☐	☐	☐	☐
pH	☐	☐	☐	☐	☐
Taste & Odor	☐	☐	☐	☐	☐
Turbidity	☐	☐	☐	☐	☐
Zeta potential	☐	☐	☐	☐	☐
Other _____	☐	☐	☐	☐	☐

Note other frequencies

C. Radioactivity

(1) Samples are analyzed each ☐ month, ☐ year, ☐ 2 years, ☐ 3 years, ☐ infrequently ☐ never. (END CARD 3) [3]

(2) Date of last radiological analysis ☐☐ mo. ☐☐ day ☐☐ yr.

(3) Analyzed by ☐ utility, ☐ state, ☐ PHS, ☐ university, ☐ other. (specify)

7. SANITARY SURVEY

A. Date of most recent survey Dy. of this data - No. of this one
(yr.) (day) (yr.)

B. Survey made by: ☐ state, ☐ PHS, ☐ local health department, ☐ utility, ☐ consultant. "Survey" requires written report of findings

C. Facilities surveyed: ☐ source, ☐ transmission, ☐ treatment, ☐ storage, ☐ distribution.

8. FACILITIES & OPERATION (describe deficiencies on reverse side)

	YES	NO
A. Are there common walls between finished and lesser quality water?	☐	☐
B. Are there inter-connections to other systems "Sub-divisions, private, industrial, community"		
(1) of known acceptable quality indicated by state health dept.		
(2) of unknown quality acceptance.		
(a) with protection Note how protected	☐	☐
(b) without protection	☐	☐
C. Is there a cross-connection control program		
(1) on new construction only	☐	☐
(2) for continuous re-inspection	☐	☐
D. Are finished water reservoirs properly covered?	☐	☐
E. Is there detectable chlorine residual in distant parts of the distribution system?	☐	☐
F. Can the treatment plant be by-passed?	☐	☐
G. Are there satisfactory procedures to:		
(1) prevent personal accidents	☐	☐
(2) prevent chlorine accidents	☐	☐
(3) disinfect all new and/or repaired distribution system mains, valves, fittings, including check samples before being placed in service?	☐	☐
H. Are there areas of low pressure (< 20 psi) in the distribution system under maximum water use?	☐	☐

I. Operating problems most often encountered are: ☐ taste & odor, ☐ phenols, ☐ corrosive water, ☐ short filter runs, ☐ other, specify.

8. FACILITIES & OPERATION, continued

J. Chlorination process was interrupted [ ][ ] times in the last 12 months.

(1) Interruptions were due to: [ ] chlorinator failure, [ ] feedwater pump, [ ] changing cylinders, [ ] power failure, [ ] other, explain.

K. Percent of land area within service area where water is available (nearest whole percent.) [ ][ ][ ] %.

L. Were plans and specs. for treatment plant approved by the state?

YES	NO
[ ]	[ ]

9. SOURCE, TREATMENT & DISTRIBUTION (Describe deficiencies on reverse side) None

A. Are the following adequate:

	YES	NO
(1) Source, with respect to the following:		
(a) quantity	[ ]	[ ]
(b) bacteriological quality	[ ]	[ ]
(c) chemical quality	[ ]	[ ]
(d) physical quality	[ ]	[ ]
(e) adequate protection	[ ]	[ ]
(2) Transmission of raw water	[ ]	[ ]
(3) Is the raw water sampled for:		
(a) Bacteriological contamination	[ ]	[ ]
(b) Chemical contamination	[ ]	[ ]
(4) Treatment, with respect to the foloving:		
(a) aeration	[ ]	[ ]
(b) chemical feed, capacity	[ ]	[ ]
(c) chemical feed, stand-by equipment	[ ]	[ ]
(d) chemical mixing	[ ]	[ ]
(e) flocculation	[ ]	[ ]

Refer to File No. 1020

9. SOURCE, TREATMENT & DISTRIBUTION, continued

A. Are the following adequate (continued):

	YES	NO
(4) Treatment, continued:		
(f) settling		
(g) recarbonation		
✓ (h) filtration At normal operating rates		
✓ (i) disinfection, capacity		
(j) disinfection, stand-by equipment		
(k) taste & odor control		
(l) fluoridation		
(5) Distribution, with respect to the following:		
(a) storage		
(b) booster chlorination		
(c) high service pumping		
(d) booster pumping		
(e) pressure		
(6) Maintenance		
(7) Records for:		
(a) disinfection		
(b) filter runs		
(c) chemical consumption		
(d) operational control tests		
(e) bacteriological examinations		
(8) Cross-connection control		
(a) ordinance		
(b) program implementation		
(c) progress		

END CARD S 5 80

9. SOURCE, TREATMENT & DISTRIBUTION, continued

Reproduced from best available copy.

B. During the past 3 years, raw water quality has ☐ 19 improved, ☐ 20 deteriorated, or ☐ 21 stayed the same.

10. PERSONNEL

A. Water Purification Operator This is the responsible plant operator - Not supt.

(1) Highest level of formal education: ☐ 22 8th grade or less, ☐ 23 high school, ☐ 24 technical or trade school, ☐ 25 university.

(2) Level of training in water treatment: ☐ college course, ☐ technical or trade school, ☐ short school, ☐ on the job, ☐ none, ☐ other, specify _____.

(3) Length of time on this job: ☐☐ years, ☐☐ months.

(4) Number of previous positions as water treatment operator ☐☐

(5) Total years of water purification experience ☐☐

(6) Level of study in sanitary microbiology: ☐ college course, ☐ technical or trade school, ☐ short school, ☐ on the job, ☐ none, ☐ other, specify. _____.

(7) Level of study in water chemistry: ☐ college course, ☐ technical or trade school, ☐ short school, ☐ on the job, ☐ none, ☐ other, specify. _____.

(8) Is the operator a full-time employee? ☐ yes ☐ no

(9) Salary range (per year) of operator: ☐ <$1,999 ☐ $2,000-4,999 ☐ $5,000-7,499 ☐ $7,500-9,999 ☐ $10,000

10. PERSONNEL, continued

A. continued

(10) Is your present staff adequate in: Operator's opinion

(a) number ☐ yes ☐ no

(b) quality ☐ yes ☐ no

B. Operator's major complaint _____

C. Most frequent customer's complaint: Received by utility.

D. Management's most frequent complaint: _____

APPENDIX D
BENEFIT/COST ANALYSIS OF CARDIOVASCULAR DISEASES AND HARDINESS

Diseases of the cardiovascular system have been responsible for approximately one million deaths annually during the past decade. In addition to causing more than half of all deaths in the United States, this heterogeneous group of diseases is responsible for a larger proportion than any other group of chronic conditions (1). Cardiovascular diseases (CVD) includes all diseases of the heart and blood vessels; they vary greatly in pathology, etiology, clinical course, and levels and trine trends of mortality.

The following subgroups, identified by the International Statistical Classification of Diseases, Injuries and Causes of Death (ICD, Seventh Revision) code numbers (2) are usually included in this broad disease category: Vascular lesions affecting the central nervous system, i.e. cerebrovascular diseases (ICD 330-334); rheumatic fever and chronic rheumatic heart disease (ICD 400-416); arteriosclerotic (including coronary heart disease) and degenerative heart disease (ICD 420-422); and other diseases of the veins (ICD 460-468). Recent changes in diagnostic standards, often make it desirable to include chronic nephritis (ICD 592-594) in the CVD category, especially if historical trends are pertinent to the analysis. A death three or four decades ago which might have been considered chronic nephritis may now be labeled hypertensive heart disease (3). In the discussion which follows the combined group including cardiovascular diseases (CVD), as defined, and renal diseases, is indicated by CV-R.

The following brief review of mortality trends since the beginning of the twentieth century describes the magnitude and significance of CV-R diseases in the United States. In 1900 the life expectancy at birth for all persons was 48 years, by 1959 it increased by more than twenty years to 69. The longevity increase for white males was less than that for white females, and the difference in life expectancy between whites and non-whites decreased significantly (4). The substantial increase to the expectancy is attributable primarily to the conquest of infectious diseases, particularly those which effected infants and young children (less than five years). Appreciable decreases in mortality due to tuberculosis and pneumonia among the middle-aged and the elderly have also contributed to the increase in the average length of life. As a consequence of this decrease in mortality, the proportion of the population 45 years and over increased from 18 percent in 1900 to 29 percent in 1960 (4). The outlook for a longer life for persons 45 years and over, relative to that for children, however, has increased only moderately. For example, a 40 year old white male in 1960 could expect to live only 3.8 years longer than a 40 year old white male in 1900. The situation is similar for other age-race-sex groups. The fundamental reason for the relatively

small increase in life expectancy for those 45 years and older has been the slow progress against cardiovascular-renal diseases, especially, atherosclerosis and until recently hypertension (4).

Although the number as well as the proportion of deaths due to CV-R appear to have increased in recent decades, it is apparent when the rates are age-adjusted (to compensate for geographical differences in the age distribution in the population) that the increase in the proportional mortality can be attributed to the increasing percentage of older persons in the population. The death rate from cardiovascular-renal diseases has actually been declining since 1930 (3). In 1967, over 54 percent of all deaths were due to CV-R diseases. In general the proportion of deaths due to CV-R increases with age; more than one-third of the deaths of persons under 65 years and over 41 percent of the mortality from all causes for the age group 35-64 years resulted from this heterogeneous category of diseases; persons in this age group are generally considered to be highly productive, and contribute more to society than they consume. Thus, in addition to their importance as a cause of death, it should be evident from the foregoing that CV-R diseases are responsible for the loss of a significant amount of manpower and economic productivity (5).

Arteriosclerosis is a condition marked by the loss of elasticity, thickening and hardening of the walls of the arteries; it may occur anywhere in the vascular system. If the inner lining of the coronary arteries is involved, the disease is called atherosclerosis and is the underlying pathology clinically manifest in coronary heart disease (1). In 1963, arteriosclerotic heart disease (ASHD; ICD 420), including coronary heart disease, was responsible for almost 550,000 deaths, approximately 55 percent of total CV-R mortality. Cerebiovascular disease (ICD 330-334) is the group of conditions typified by vascular lesions affecting the central nervous system (1); it is commonly referred to as stroke. These diseases claimed over 200,000 lives, or 20 percent of all CV-R mortality in 1963. These two disease categories (ICD 330-334 and ICD 420) accounted for more than three-fourths of all CV-R disease deaths.

A National Health Survey conducted in 1960-1962 further illustrates the significance of cardiovascular diseases in the United States. The prevalence of several entities of heart disease was extrapolated from extensive examinations of almost 6700 persons between 18-79 years. The investigation revealed that over 27.5 million persons either definitely had (14.6 million) or were suspected of having (12.9 million) some form of heart disease. In other words, 132 of every 1000 persons in this age group exhibited definite heart disease among the specific components of heart disease considered in the survey, hypertensive heart disease (HHD) was the most frequently encountered. Over ten million American adults had definite signs of HHD, and an additional five million were categorized as suspects. Coronary heart disease was the second most frequent disease observed with over three million definite and nearly two and a half million suspected cases (3).

All of the disease entities except AHD, which are categorized as CV-R diseases have exhibited a down trend similar to that previously noted for CV-R in general. Arteriosclerotic heart disease increased by 23 percent, from 248 per 100,000 to 308 per 100,000, for white males during the 1949-1962 period. The increase for non-white males during the same period was over 50 percent, from 150 per 100,000 to 235 per 100,000. Age adjusted AHD mortality rates for women also increased during the 13 year period; the increase, however, was significantly lower than that exhibited by men (3). Although some of the observed increases can be attributed to changes in both the definition of AHD and the method of diagnosis, it is felt that the trends are in fact due to a real increase in the incidence of AHD. In an extensive review of coronary heart disease, Epstein (6) concluded that almost no progress has been made in the control of CHD since 1950.

As a direct consequence of the significant impact cardiovascular diseases, in general, and arteriosclerosis, in particular, have on society and the economy (5), there have been numerous epidemiological investigations. Epidemiology is properly regarded as a field of activity which relates population studies of a specific disease to clinical and laboratory investigations (6). Significant variation in AHD mortality rates from one country to another has been recognized for many years. Generally mortality rates appear to vary directly with the standard of living, yet there is extreme variation among those countries with similar standards of living and medical care; the United States, Canada and Australia have higher AHD mortality rates than the Netherlands, Sweden and Switzerland (6). Although it is apparent that the high mortality rate for AHD experienced in the U.S. is not shared by all highly developed nations, it is important to recognize the existence of geographical variation in the incidence of this disease entity (5). Geographical differences in CV-R mortality also prevails within the United States, and geographic clustering with respect to CV-R mortality rates is evident. A study by Sauer (7) using the CV-R mortality rates for white males age 45-64 from 509 state economic areas, indicated that the highest rates were near the East Coast and the lowest rates occurred in the West Central and Mountain regions. According to Sauer, if the lowest mortality rates were applied to the total U.S. population, there would be between 150,000 and 200,000 fewer deaths annually. Analogous trends can also be shown for white females CV-R mortality, as well as for white male and female mortality from arteiosicerotic heart disease and hypertensive heart disease. Patterns for non-whites are more difficult to discern due to the distribution and population density of this group. Enterline, et.al. (8) compared 1950 CHD death rates from metropolitan with non-metropolitan areas. Although they found that CHD mortality was more than 40 percent higher in metropolitan counties which contained a central city than in non-metropolitan counties, they concluded that state to state difference cannot be accounted for solely by the degree of urbanization. Sauer and Enterline (9) analyzed the geographic variation (by state) in cardiovascular mortality. They found the differences to be real, and of considerable magnitude in some instances, and not dependent upon the quality and

quantity of physicians available; it was also determined that the population and death counts were fairly accurate, as were the cause-of-death entries on death certificates.

The identification of mortality and morbidity patterns such as the geographic ones discussed above have greatly assisted epidemiologist in their investigations. Attributes which appear to occur more commonly among persons with CVD than in a control population are designated as "risk factors" (6). Analysis of a large number of attributes of a population under study and statistical evaluation have permitted the identification of several factors which conform to the criteria necessary to be specified as "risk factors." Serum triglycerides and cholesterol levels, blood pressure levels, smoking, obesity, physical activity, psychological and social factors, such as stress, genetic factors and diet are the most frequently cited with respect to various entities of heart disease (3). Some studies have found variables, such as air pollution, climate, water hardness and soil composition, that describe the physical and chemical nature of the environment to be significantly correlated with CVD, thus to the risk of a premature death (6, 7, 10-15).

Notable geographical differences in the incidence of death from apoplexy (cerebral hemorrhage) led Kabayaski (16) to consider the role of environmental factors. He found that the river water in those areas which exhibited high apoplexy mortality rates have an excessively high sulfate to carbonate ratio. He concluded that the correlation between mortality due to apoplexy and the chemical composition of the water was a result of the abnormally acidic nature of many Japanese rivers, a function of the geological construction of the catchment area. Schroeder (17) investigated mortality from all heart disease in the same areas of Japan and found an even higher correlation between this parameter of water quality and mortality. Subsequently Schroeder examined the observed geographic variation in U.S. cardiovascular mortality and several parameters used to describe the chemical and physical characteristics of potable water (13). He found highly significant negative correlations, i.e., high levels of the water parameter were associated with lower mortality, between cardiovascular mortality and magnesium, calcium, bicarbonate, sulfate, fluoride, dissolved solids, specific conductance and pH. The most significant correlations were between hardness and main constituents, calcium and magnesium, and CVD. He concluded that, "Some factor either present in hard water or missing or entering soft water appears to affect death rates from degenerative cardiovascular diseases."

These finds stimulated similar investigations in several other countries during the next five years. In 1961 Morris, et.al. (18), reported a highly significant correlation between water hardness and CVD mortality in 83 county boroughs in England and Wales. The correlation coefficients were generally in remarkable agreement with those obtained earlier by Schroeder. The highest correlation (-0.65 at the 0.001 probability level) was found between males 45-64 years, and calcium. A notable difference between the British and the U.S. studies was the absence a correlation between magnesium and mortality in the former study. Muss (19) investigated CVD

mortality for the years 1936-1940 for health districts in New York City served by two different water sources, one of which was soft, and the other relatively hard. He reported that the mortality rate was ten percent lower in the hard water districts. In Sweden, Biorck, et.al. (20), investigated the rates due to several important causes of death for 34 cities with respect to geology, water source and 23 chemical parameters of drinking water. Highly significant negative correlations ($p < 0.005$) between male (25-44 years) mortality due to degenerative disease of the heart (ICD 422) and water hardness, between male (25-64 years) mortality due to ICD 422 and calcium and between female (65-74 years) mortality resulting from cerebrovascular was correlated with total water hardness at the 0.01 level of significance. In general, correlations which involved magnesium or AHD (ICD 420) for any age-sex group were not signifcant. A later study by Bostrom and Wester considered mortality in Sweden's three largest cities. They reported that hard water was related to decreased cardiovascular mortality among women but not men.

Anderson and co-workers (22) and Anderson and LeRiche (23) found that death rates from ischemic heart disease (ICD 420) were higher in the soft water areas of Ontario, Canada, than in the hardwater areas. Deaths were classified as being sudden or non-sudden (on the basis of whether or not the death certificate was signed by a coroner). They concluded that the excess mortality in the softwater areas relative to the hardwater areas could be attributed to the higher incidence of sudden death in the soft water areas. An investigation by Peterson, et. al., in the U.S., supports these finds (24). It has been suggested that the high correlation between ischemic heart disease and water hardness may be the result of increased susceptibility to fatal arrythmics associated with myocardial infraction among residents of soft water areas. Neri, et.al. (25), also examined the relationship between water hardness and sudden death and suggested the possible significance of trace metals.

Marsironi (26) considered the death rates in two hard-water and two soft-water river basins. He found that the alpah radio-activity was inversely correlated with mortality rates in the study area: the correlation coefficient between alpha radiation and water hardness was found to be 0.86. Mortality due to hypertensive heart disease and ischemic heart disease was lower by 41 percent and 25 percent, respectively in the hard water river basins than in the soft water river basins.

Several studies have examined populations that have been subjected to substantial changes in the hardness of their water supply (27-29). The finding of these investigators support the inverse relationship between CVD mortality and water hardness. Crawford, et.al., found that the change in CVD mortality in those county boroughs which changed from hard water to soft water was twice that experienced by county boroughs that did not change their source supply (29).

Schroeder suggested that trace metals in the water, either initially present or leached from the distribution system may be an importnat factor in cardiovascular mortality (3). The corrosiveness of water was also cited as a potentially important factor. He proposed that cadmium, a known factor in high blood pressure, may be picked up from galvanized pipes in the home distribution system. Bostrom and Wester (21) and Sauer (31) investigated the trace element content of drinking water in Sweden and the United States, respectively. Both studies were unable to establish a significant correlation between CVD mortality and the concentrations of various trace elements.

Schroeder and Kraemer found highly significant inverse correlations for 15 constituents of water (12 were metals) with cardiovascular mortality (32). They also examined the corrosiveness of water, as measured by Langlier's index, and found direct correlation with atherosclerotic heart disease at high levels of significance. This is in agreement with Kogayashi's conclusions seventeen years earlier in Japan (16). Schroeder and Kraemer point out, in view of their statistical analysis, that the chemical composition of tap water may be a more important etiologic factor than that of finished water (at the treatment facility).

Although the studies discussed thus far, in addition to several other investigations (33037) generally confirm the existence of an inverse correlation between CVD mortality and a 'water factor,' there have been several epidemiological studies which do not support the relationship. Mulcahy examined the 15 largest urban centers in Ireland; he was unable to establish any significant correlation between hardness and heart disease (38). In another study of the 20 largest urban areas in the Republic of Ireland and Northern Ireland, he found a consistent inverse relationship, however, it was not statistically significant (39). An investigation by Lindeman and Assenzo in Oklahoma also failed to establish a significant inverse relationship (40). Strong, et. al., as part of the International Atheriosclerosis Project, investigated the possible association between the mineral content of drinking water and atherosclerotic lesions in the aorta and coronary arteries. Clinical examination of samples from 14 countries revealed an insignificant negative association between water hardness and the prevalence and extent of atherosclerosis (41).

It has been suggested that the data used in Oklahoma and Ireland were biased (29); the death rates were based upon very small populations, frequently from small towns or rural areas; the range of water hardness in the Irish study was small and in Oklahoma, Lindeman and Assenzo failed to weight the hardness of the ponulation served. The validity of the international investigation can be questioned due to the small number of areas considered as well as the lack of sample homogeneity.

Statistical correlations between some constituents of water and various entities of cardiovascular disease have been established in several areas of the world. It is highly unlikely that these relationships are simply coincidental or spurious however, substantial and reasonable skepticism exists because pathological explanations are not yet apparent.

Some investigators have considered several socioeconomic, climatic, geologic and other environmental factors in order to identify variables which may be significantly related to CVD and to establish mutual associations between certain factors, i.e. water composition and socioeconomic factors, and mortality (11, 12, 14, 18, 31, 35, 42, 43). Crawford, et. al. (33) found rainfall to be highly associated with calcium. This relationship is reasonable since areas of high rainfall generally coincide with soft water areas; water supplies in these areas are in contact with the geologic formations for shorter periods of time. A multiple regression analysis indicated that both calcium and rainfall made significant contributions ... to the variation in CVD mortality after allowing for the effect of other environmental variables (44). Dudley, et. al., found that the comfort index, was significantly correlated with CVD mortality (14). The authors suggested this derived climatological variable may actually be representative of physiological stress due to the environment. An analysis by Sauer which considered 37 socioeconomic variables found that cultural and geographic factors were consistently important in explaining the death rates exhibited by various entities of CVR (31). Mobility, employment, income and education were found to be significantly correlated with mortality. Lehr, et.al., investigated the relative importance of 12 social variables and 12 biological risk factors by applying a discriminant analysis technique (42). They concluded that the religion of the parents, education and father's income were the most significant social variables.

Despite the significant statistical relationships that have been established by many investigators in several countries, little progress has been made in identifying the possible mechanisms in pathogensis or factors in drinking water by investigating the clinical and biochemical characteristics of populations with differing water supplies. There are four studies worthy of mention which have investigated some important pathological and physiological aspects of the water-CVD relationship.

Stitt, et.al., compared several clinical and biochemical measurements of males from six hardwater and six softwater towns in England and Wales (45). They found that the mean values for blood pressure, plasma-cholesterol, and heart rate were higher in the softwater group. Bierenbaum, et.al., examined the serum lipids of two groups of men 33 to 39 years; one group resided in a softwater area, the other in a hardwater area (46). The serum levels of calcium, magnesium, cholesterol and triglycerides were all higher in the hard water location. A conclusion concerning CVD mortality is difficult to make due to the small size of the sample. Crawford and Crawford conducted a microscopic investigation of the three primary blood vessels of the heart in order to determine differences in the extend of cardiac lesions in a soft water and hard water area (47). The authors suggested that the myocardium of persons living in soft water areas may be more susceptible to fatal arrythmicis. In another study by the same authors, it was found that the bone content of lead in soft water areas was significantly higher than it was in hard

water areas (48). They concluded that the role of lead in CVD could not, at that time be answered, however, they suggested that lead may be replacing calcium and magnesium in some important relationships.

The preceding summary of mortality trends and current epidemiological investigations highlights those aspects of cardiovascular diseases which are relevant to the present study. The objective of this investigation is to develop a model which relates CVD to water quality, and can be used in a benefit/cost analysis for the evaluation of investments in water supply programs.

Although it has been evident for sometime that CVD mortality is a function of the interaction of many factors, including the physical and social environment, most of the previously discussed examinations which relate mortality from diseases of the cardiovascular system with water quality, failed to consider the role of other environmental factors. While this approach is informative during initial investigations, it does have limited value; it should be noted that relationships designated as significant by simple correlation analysis are often misleading and may result in misinterpretation of the data. Sauer (31) and Gardner co-workers (44) by the application of multiple regression analysis, have attempted to establish the role of water in CVD mortality in a more realistic fasnion. The literature indicates that in addition to variables which characterize the chemical composition and quality of the water supply, it is necessary to include variables that describe the socioeconomic, climatological and physical environment of the population being studied (Sauer, 31; Crawford, et.al., 33; Gradner, et.al., 44; Dudley, et.al, 1969; Lehr, et.al., 42). Although the inclusion of risk factors (smoking habits, diet, obesity, serum triglyceride cholesterol levels, and blood pressure, etc.) as independent variables is highly desirable, the nature of this information obviates its use in the present analysis. Temporal and financial limitations did not permit the acquisition of original data, thus the investigation was restricted to the use of previously collected information. This sometimes results in formidable incongruities in the data, however in the current investigation these inconsistencies were not considered to have a significant effect on the results.

The data required to conduct the present study can be grouped into five categories: 1) water composition and quality; 2) climatological; 3) socioeconomic and population; 4) mortality; and 5) costs of CVD in the United States. The study of Durfor and Becker (49) on the Public water supplies of the 100 largest cities in the United States provides the most recent and best available data on the chemical composition and quality of water publically consumed. The climatological data were extracted from monthly U.S. Weather Bureau reports (1959-1961). Duffey and Carroll (51) compiled mortality data by age, sex and color for all major causes of death in 212 Standard Metropolitan Statistical Areas (SMSAS) for 1959-61. Socioeconomic and population data were taken from the 1962 County City Data Book and the 1960 Census of Population, respectively. Two thorough investigations (1962, 1967) which analyzed the

economic impact of CVD were used to determine the benefits which would accrue as a result of a decrease in cardiovascular disease mortality.

The present study was restricted to the consideration of metropolitan populations, as a consequence of data availability. This situation was further complicated since on one hand the population was defined by the water distribution system, and on the other by economic and administrative boundaries which define the SMSA's; it was difficult to ascertain how well the data taken from the water distribution population represented the same characteristics of the SMSA population and vice-versa. Thus it was necessary to develop a criteria which could be used to evaluate the potential significance of differences between the two populations. Durfor and Becker (49) provided information on the percent of finished water which was supplied by each source. It was decided that if the product of the percent of supply and the ratio of the population served to the population of the SMSA was less than 63 percent the city would be excluded from the sample. In this manner cities for which the socio-economic and mortality data, and water quality data were not representative of the same population were eliminated from the statistical analysis. Fifty-six cities representing fifty-three SMSA's met the stated criteria (see Table 1); the total population was nearly 51 million. For those cities supplied by more than one source, the weighted average, based on the relative proportion of finished water from each source was used to determine the concentration of each chemical constituent. Thus, it was assumed that all waters were mixed prior to distribution, that is, composition was assumed to be uniform throughout the system. Although this was not true for systems which did not combine their finished water prior to distribution, the assumption is reasonable if the percent of supply is considered equivalent to percent of population served. Data from regular determination were used in preference to the analytical results from a one day sample, since the former were the average values from several determinations, thus more representative of the actual concentrations. The parameters used to describe the chemical composition and quality of municipal waters are listed in Table 2.

The annual mortality rates used in this investigation were derived from a three-year average (1959-1961). The rates were expressed as the number of deaths per 100,000 population, and age-adjusted by the direct method using 10-year age groups (Duffey and Carroll, 51). This adjustment compensates for differences in the age distribution of the population among the cities. Three disease categories were chosen as the dependent variables for the statistical analysis: 1) mortality from all cardiovascular diseases (ICD 330-334, 400-468); 2) mortality from heart diseases of obscure etiology (Arteriosclerotic heart disease [ICD 420], hypertensive heart disease [ICD 443-447] and general arteriosclerosis [ICD 450]); and 3) arteriosclerotic heart disease (ICD 420). Four population subsets were chosen for each disease category: 1) total population; 2) all white males; 3) white males, 35-74 years; and 4) white males, 45-64 years. Thus, twelve dependent variables were considered in the development of a general model; these twelve variables are highly correlated (probability level less than 0.0001) with each other (see Table 3). Several highly

TABLE 1. Standard Metropolitan Statistical Areas

Alabama:
- Birmingham
- Mobile
- Montgomery

Arizona:
- Tucson

California:
- San Francisco-Oakland

Colorado:
- Denver

Connecticut:
- Bridgeport
- Hartford
- New Haven

Florida:
- Tampa-St. Petersburg

Georgia:
- Savannah

Illinois:
- Chicago
- Rockford

Indiana:
- Evansville
- Fort Wayne
- Indianapolis

Iowa:
- Des Moines

Kansas:
- Wichita
- Topeka

Kentucky:
- Louisville

Louisana:
- Baton Rouge
- New Orleans

Maryland:
- Baltimore

Massachusetts:
- Boston

Michigan:
- Detroit

Mississippi:
- Jackson

Nebraska:
- Lincoln
- Omaha

New Mexico:
- Albuquerque

New York:
- New York City-Yonkers

North Carolina:
- Charlotte

Ohio:
- Cincinnati
- Cleveland
- Columbus
- Toledo

Oklahoma:
- Oklahoma City
- Tulsa

Oregon:
- Portland

Pennsylvania:
- Erie

Tennessee:
- Chattanooga
- Memphis
- Nashville

Texas:
- Amarillo
- Austin
- Corpus Christi
- Dallas
- El Paso
- Lubbock
- San Antonio

TABLE 1. Standard Metropolitan Statistical Areas (Continued)

Virginia:
- Norfolk
- Richmond

Washington:
- Seattle

Wisconsin:
- Milwaukee

TABLE 2. Chemical Composition and Quality of Water. (all concentrations were parts per million, except for Chromium, Copper, Lithium and Manganese which were parts per billion)

Calcium (53)*

Chromium (42)**

Copper (51)

Lime-Soda Softening (53)

Lithium (51)

Magnesium (53)

Manganese (44)**

Percent of Supply (53)

Potassium (52)

Silicon Dioxide (53)

Sodium (52)

Source of Supply (53)

Total Dissolved Solids

Raw Water (38)**

Finished Water (53)

Total Hardness

Raw Water (38)**

Finished Water (53)

* Numbers in parentheses indicate the number of SMSA's for which data were available.

** Exclude from statistical analysis, sample size less than 50.

TABLE 3. Product Moment Correlation Coefficients for Dependent Variables

	TOTCVD	CVDWHM	CVD3574	CVD4564	TOTHD	HDWHM	HD3574	HD4564	TOTAHD	AHDWHM	AHD3574	AHD4564
TOTCVD	1.00	0.96	0.84	0.84	0.85	0.84	0.79	0.81	0.65	0.80	0.76	0.78
CVDWHM		1.00	0.83	0.83	0.88	0.90	0.81	0.81	0.72	0.85	0.78	0.79
CVD3574			1.00	0.95	0.79	0.78	0.93	0.91	0.68	0.77	0.91	0.89
CVD4564				1.00	0.78	0.78	0.91	0.96	0.66	0.75	0.89	0.94
TOTHD					1.00	0.98	0.90	0.86	0.91	0.93	0.87	0.85
HDWHM						1.00	0.90	0.87	0.90	0.97	0.88	0.86
HD3574							1.00	0.96	0.83	0.89	0.99	0.95
HD4564								1.00	0.78	0.85	0.95	0.99
TOTAHD									1.00			
AHDWHM									1.00	0.89	0.85	0.80
AHD3574										1.00	0.90	0.86
AHD4564											1.00	0.96
												1.00

TOTCVD Total cardiovascular diseases
CVDWHM Cardiovascular diseases, white males
CVD3574 Cardiovascular diseases, white males, 35-74 years
CVD4564 Cardiovascular diseases, white males, 45-64 years
TOTHD, HDWHM, HD3574, HD4564 as above except mortality due to arteriosclerotic heart disease, hypertensive heart disease and general arteriosclerosis
TOTAHD, AHDWHM, AHD3574, AHD4564 as above except mortality due to arteriosclerotic heart disease.

correlated dependent variables were used in order to distinguish between meaningful and spurious correlations when examining the relationship between the dependent and independent variables. It was felt that those independent variables which correlated significantly with several dependent variables would be indicative of a non-spurious correlation between the independent variables and CVD. Cardiovascular disease mortality among white males 45-64 years, is of considerable interest due to the relatively high death rate exhibited by this segement of the population, and also due to the significant economic loss of society resulting from the premature death of persons during their height of productivity. The cause of death for persons in this age group can generally be attributed to specific factors. Thus, the data are more reliable and amenale to epidemiological investigations. Non-white mortality data have been excluded since the proportion of the metropolitan populations due to this group was relatively small.

Many of the climatological variables used in this investigation were previously studied by Dudley, et.al., (14). Average measurements for three successive years (1959-1961) were used to compensate for possible extreme conditions for any one year, and also to ensure agreement with the mortality data. Average values for the months of January and July were used, since these months generally represent the annual climatic extremes, thus the maximum stress induced by the climate. The variables used can be classified as those measured directly, or those derived from direct measurements (see Table 4). Values for all of the socioeconomic variables, except the percent employed in agriculture, were taken from the 1962 County-City Data Book (see Table 5). The percent employed in agriculture is simply,

$$\frac{\text{number employed in agriculture}}{\text{total labor force}} \times 100$$

Multiple regression analysis was selected as the most appropriate and expedient means of establishing the relative importance of several variables with respect to their role in explaining the variation in the dependent variable (mortality). Due to the large number of independent variables, it was desirable to reduce their number by some meaningful and rational procedure prior to the application of multiple regression. All statistical analyses were performed by the use of the Statistical Analysis System prepared by Barr and Goodnight (54).

The model was developed according to the procedure described in the following discussion. Product moment correlations were computed for each possible pair of independent and dependent variables. Those independent variables which met the criteria of being correlated with each of the twelve dependent variables at the 0.05 level of sigaificance were selected for use in a stepwise regression. All other independent variables were excluded from further consideration. This procedure reduced the number of independent variables from over 50 to 16, and hopefully in addition, eliminated spurious correlations. The stepwise technique selects that combination of independent variables which produces the largest multiple

TABLE 4. Climatological Variables

Direct Measurements

Altitude (ALT), feet
Latitude (LAT),
Average Temperature January (AVT1), °F
Average Temperature July (AVT7), °F
Precipitation January (PPT1), inches
Precipitation July (PPT7), inches
Annual Precipitation (PPTYR), inches
Relative Humidity January (RH1), percent
Relative Humidity July, (RH7), percent
Atmospheric Pressure July (ATM7), millibars

Derived Measurements

Temperature Variation January (TV1), °F
Temperature Variation July (TV7), °F
(Temperature Variation = Average monthly maximum - Average monthly minimum)
Annual Temperature Variation (ATV), °F
(ATV = AVT7 - AVT1)
Comfort Index January (CI1)
Comfort Index July (CI7)
(CI = 0,55 [1-RH] [AVT - 58])
Combined Comfort Index (CCI)
(CCI = $|CI7| + |CI1|$)
Temperature Humidity Index January (THI1)
Temperature Humidity Index July (THI7)
(THI = AVT - CI)

TABLE 5. Socioeconomic Variables

Median income, hundreds of dollars (MEDINC)

Median number of years of school completed (MEDSCH)

Percent employed in agriculture (PAGEMP)

Percent finished high school (PFINHS)

Percent living in one unit structures (P1UNIT)

Percent urban (PURB)

Percent white collarworkers (PWHTCOL)

Percent owning home freezers (PFREEZ)

Percent owning two or more automobiles (P2AUTO)

Percent with income greater than $10,000 (INCGT10)

Percent with income less than $3,000 (INCLT3)

Percent with more than one person per room (PGE1PRM)

Persons per housing unit (POPUNIT)

Population density of central city (CORDEN)

correlation coefficient. Each of the selected variables must also be significant at a specified level, that is, the contribution of each in explaining the variation of the dependent variables must satisfy predetermined conditions of experimental design. The twelve equations (one for each dependent variable) produced by this procedure were examined to determine patterns as well as the frequency of occurrence of the variables in the twelve models. In this manner the most important variables were selected for the final step in developing a general model. Meaningful combinations of these variables were used in multiple regression analysis and the model which most frequently exhibited the highest R^2 (percent of variation explained) was selected as the general model.

Although other combinations of independent variables resulted in higher R^2s, the improvement was generally insignificant. The variables included in the final model should not be considered to be causative agents, although they may be representative of/or related to factors which have a causative role in CVD mortality.

Table 6 contains the correlation coefficients of the independent variables which were found to correlate ($p < 0.05$) with all 12 dependent variables. All of the correlations were inverse, except for those between the population density of the central city (CORDEN) or temperature variation July (TV7) or altitude (ALT) and any dependent variable. The inverse correlation between mortality and the variables which describe the chemical composition of the water are in agreement with Sauer's (37) findings, as the concentration of these substances increased mortality due to CVD decreased. Table 7 shows the mean concentrations of these constituents for the 53 MSA's ranked by total cardiovascular mortality rate per 100,000 (TITCVD).

Table 7

	TOTCVD	SIO2	NA	LI	TDS
Lowest Rate	426.7	15	61	13	283
	501.7	8	13	8	164
Highest Rate	572.8	7	1	2	128
Mean Values	499.0	11	2	8	191

All of the variables designated by a double asterisk (**) in Table 6 were found to have at least six of the twelve correlations significant at the 0.0005 level, and the percent of families owning two or more automobiles (P2AUTO) was correlated at this highly significant level with all twelve dependent variables. Gardner, et.al. (44), found a similar variable (number of households per automobile to be significantly correlated with mortality and derived variable [social factor score]) which purports to measure socioeconomic status.

TABLE 6. Correlation (r)* of 12 Categories of Mortality with Water, Socioeconomic and Climatological Variables

	TOTCVD	CVDWHM	CVD3574	CVD4564	TOTHD	HDWHM	HD3574	HD4564	TOTAHD	AHDWHM	AHD3574	AHD4564
SIO2**	-0.43	-0.44	-0.38	-0.34	-0.52	-0.55	-0.50	-0.41	-0.48	-0.53	-0.49	-0.42
NA	-0.42	-0.46	-0.40	-0.38	-0.45	-0.47	-0.40	-0.35	-0.43	-0.46	-0.39	-0.35
K	-0.47	-0.43	-0.34	-0.36	-0.50	-0.51	-0.43	-0.39	-0.43	-0.51	-0.43	-0.39
TDS**	-0.49	-0.48	-0.44	-0.44	-0.49	-0.50	-0.48	-0.44	-0.43	-0.50	-0.49	-0.46
LI	-0.39	-0.40	-0.32	-0.36	-0.40	-0.43	-0.37	-0.35	-0.36	-0.45	-0.38	-0.37
CORDEN	0.30	0.33	0.36	0.30	0.51	0.47	0.45	0.34	0.62	0.49	0.45	0.34
MEDSCH	-0.31	-0.45	-0.36	-0.36	-0.40	-0.38	-0.31	-0.33	-0.27	-0.35	-0.28	-0.30
PFINHS	-0.57	-0.51	-0.39	-0.39	-0.45	-0.44	-0.35	-0.36	-0.30	-0.41	-0.33	-0.34
PFREE2	-0.31	-0.28	-0.41	-0.44	-0.47	-0.42	-0.49	-0.48	-0.50	-0.42	-0.48	-0.47
P2AUTO**	-0.57	-0.55	-0.59	-0.55	-0.56	-0.56	-0.57	-0.52	-0.52	-0.52	-0.53	-0.50
P1UNIT**	-0.34	-0.35	-0.42	-0.35	-0.52	-0.49	-0.49	-0.38	-0.64	-0.52	-0.50	-0.38
ALT**	-0.51	-0.54	-0.54	-0.50	-0.51	-0.57	-0.57	-0.52	-0.44	-0.57	-0.56	-0.52
TV7	-0.38	-0.39	-0.30	-0.31	-0.43	-0.45	-0.37	-0.35	-0.43	-0.49	-0.38	-0.36
ATM7**	0.45	0.51	0.46	0.46	0.49	0.55	0.51	0.48	0.45	0.56	0.52	0.50
RH7**	0.55	0.55	0.48	0.49	0.47	0.52	0.49	0.49	0.32	0.51	0.48	0.49
CI7	-0.49	-0.53	-0.40	-0.39	-0.44	-0.50	-0.41	-0.36	-0.38	-0.47	-0.40	-0.37

* Significance levels of the product moment correlation coefficient (r): $r < -0.26$ or $r > 0.26$: $p < 0.05$
$r < -0.39$ $r > 0.39$: $p < 0.005$
$r < -0.48$ $r > 0.48$ $p < 0.0005$

** At least six of the twelve correlations are significant at th 0.0005 probability level.

Table 8 lists all additional independent variables that appear to be significantly correlated with the dependent variables, however, they did not satisfy the criteria being correlated at the 0.05 probability level with all twelve dependent variables. Table 9 shows that each of these six variables were significantly correlated with at least one variable in Table 6, thus justification for the exclusion of these variables from further consideration. (This action was later verified by considering a stepwise regression with all 22 independent variables; the results from this stepwise regression were similar to those using the 16 variables in Table 6, as discussed below).

Table 9. Correlation (r) of Variables in Table 8 with Variables in Table 6

Table 8	Table 6	Correction Coefficient	Probability Level
CA	TDS	0.77	0.0001
PWHTCOL	PFINHS	0.82	0.0001
PAGEMP	P1UNIT	-0.47	0.0007
PPTI	ALT	-0.52	0.0002
PPTYR	ALT	-0.66	0.0001
TVI	ALT	-0.62	0.0001

The 16 independent variables in Table 6 were entered one by one, as discussed previously, into a stepwise regression analysis; the probability level for inclusion in the model was chosen to be 0.05. Inspection of the twelve regression equations from this analysis indicated the predominance of certain variables. In general, at least one variable from each of the independent variable categories (water, socioeconomic and climatological) appeared in each regression equation. Table 10 shows those variables which appeared most frequently, in addition to the correlation coefficients between those variables within each category.

Based upon the results of the stepwise regressions, the high correlation between the variables within each category, and the frequency of appearance in the twelve regression models, five variables (TDS, K, P2AUTO, PFREEZ, ATM7) were retained for further analysis. From the foregoing discussion, it is evident that the final model will have three independent variables, one from each category. The four possible combinations of three variables (one variable from each group) were used in a multiple regression analysis. The variable combination of TDS, P2AUTO, and ATM7 resulted in the highest R^2 (percent of variation explained) for seven of the twelve regression equations. TDS, PFREEZ and ATM7 resulted in the higest R^2 for the other five equations. The average R^2 for the twelve models containing TDS, ATM7 and either P2AUTO or PFREEZ was 0.46 and 0.43, respectively. Thus, on the average the model,

$$\text{CVD MORTALITY} = \beta_0 + \beta_1 \text{ TDS} + \beta_2 \text{ P2AUTO} + \beta_3 \text{ ATM7}$$

TABLE 8. Correlation (r)* of 12 Categories of Mortality with Six Independent Variables

	TOTCVD	CVDWHM	CVD3574	CVD4564	TOTHD	HDWHM	HD3574	HD4564	TOTAHD	AHDWHM	AHD3574	AHD4564
CA	-0.32	-0.26	-0.28	-0.28	-0.30	-0.30	-0.34	-0.31	----	-0.31	-0.35	-0.32
PWHTCOL	-0.53	-0.50	-0.37	-0.38	-0.38	-0.29	-0.31	-0.34	----	-0.36	-0.30	-0.33
PAGEMP	-0.44	-0.36	-0.28	---	-0.47	-0.40	-0.35	-0.30	-0.45	-0.41	-0.34	-0.29
PPTI	0.53	0.53	0.44	0.41	0.42	0.46	0.39	0.36	----	0.43	0.37	0.34
PPTYR	0.54	0.51	0.48	0.48	0.42	0.45	0.46	0.46	----	0.45	0.46	0.46
JVI	-0.29	-0.37	-0.28	---	-0.47	-0.49	-0.38	-0.30	-0.55	-0.48	-0.37	-0.30

* See Table 6 for explanation of significance levels.

TABLE 10. Correlation* Between Variables Within Each Category.

Water Composition

	NA	K	TDS	
TDS	0.80	0.77	1.00	(3)**
K	0.67	1.00		(3)
NA	1.00			(1)

Socioeconomic

	PFREEZ	P1UNIT	P2AUTO	
P2AUTO	0.53	0.65	1.00	(6)
P1UNIT	0.57	1.00		(2)
PFREEZ	1.00			(3)

Climatological

	ATM7	RH7	
RH7	0.78	1.00	(2)
ATM7	1.00		(10)

* See Table 6 for explanation of significance levels.

** The numbers in parentheses indicate the frequency that each variable represented its respective group in the regression equations.

were β_0 is the intercept, and β_1, β_2 and β_3 are the coefficients of regression, explained a higher proportion (46%) of variation in the twelve dependent variables than the variable combination in which PFREEZ was substituted for P2AUTO. The coefficients (β - values) and pertinent statistical information for the twelve models are given in Table 11.

Examination of the equation which describes the relationship between total cardiovascular mortality and water quality (equation 1, Table 11) should be beneficial in understanding the general model.

TOTCVD = 442.07 - 0.14 TDS -5.30 P2AUTP + 0.21 ATM7

If P2AUTO and ATM7 are held constant while TDS is varied, the change in TOTCVD can be expressed as follows:

$$\Delta TOTCVD = -0.14\ \Delta TDS$$

A ten part per million increase in total dissolved solids, according to the model, would result in a decrease in mortality of 1.4 per 100,000.

If all the SMSA's were provided with water containing a total dissolved solids concentration equivalent to that of those cities ranked in the lowest third by mortality, the mean concentration would be increased by 92 ppm, from 191 ppm to 283 ppm (See Table 7). The expected decrease in mortality due to an increase in TDS is 12.9 per 100,000, a 2.6 percent decrease in the death rate. Thus in 1960, approximately 6600 fewer persons would have died from CVD in the 53 SMSA's considered in this study. Similarly, it can be shown that a 92 ppm increase in TDS would have resulted in decrease in CVD mortality of 30.4 per 100,000 categorized as white males (45-64 years), again a 2.6 percent reduction. The economic impact of this loss of life can be assessed if the present values of future earnings at the time of death (1960) were known for these persons. For illustrative purposes, a crude, but reasonable estimate of $34,000 can be used. A 2.6 percent reduction in mortality is equivalent to saving 1400 lives in this age-sex-race group. The economic benefits resulting from the consideration of future earnings alone is approximately $48 million. Additional monetary benefits due to decreased medical expenditures and losses due to morbidity can also be calculated.

In general, the dissolved solid content of finished water is lower that that of raw water. Water softening processes are responsible for the largest reduction in TDS, although other treatment processes, such as filtration and coagulation, also remove significant amounts of dissolved substances. The removal of minerals prevents the formation of deposits in the distribution system, boilers and hot water heaters (economic benefits). Softening also provides water which requires the use of less soap and detergent, and is aesthetically pleasing. Thus it is difficult to justify increasing the dissolved solids concentration in water.

TABLE 11. Models Which Describe the Relationship Between Mortality From Cardiovascular Diseases and Variables From the Physical and Socioeconomic Environment.

Mortality Category (Rate/100,000)	Intercept	TDS*	P2AUTO	ATM7	R^2
All Cardiovascular Diseases:					
1) Total Population, TOTCVD	442.07	-0.14 (0.0375)	-5.30 (0.0005)	0.21 (0.0992)	0.47
2) All White Males, CVDWHM	445.99	-0.11 (0.0851)	-4.82 (0.0009)	0.30 (0.0203)	0.47
3) White Males, 35-74 years, CVD3574	2831.00	-0.70 (0.1437)	-40.04 (0.0003)	1.83 (0.0495)	0.46
4) White Males, 45-64 years, CVD4564	1238.26	-0.36 (0.1256)	-17.58 (0.0009)	0.87 (0.0601)	0.43
Heart Disease of Obscure Etiology:					
5) Total Population, TOTHD	222.53	-0.10 (0.0572)	-4.09 (0.0007)	0.22 (0.0343)	0.48
6) All White Males, HDWHM	228.88	-0.11 (0.0709)	-4.42 (0.0009)	0.33 (0.0061)	0.51
7) White Males, 35-74 years, HD3574	1674.77	-0.75 (0.0796)	-33.43 (0.0006)	1.98 (0.0186)	0.49
8) White Males, 45-64 years, HD4564	745.67	-0.33 (0.1413)	-14.80 (0.0025)	0.96 (0.0304)	0.42
Arteriosclerotic Heart Disease:					
9) Total Population, TOTAHD	177.08	-0.07 (0.1528)	-3.15 (0.0025)	0.18 (0.0596)	0.39
10) All White Males, AHDWHM	169.45	-0.10 (0.0638)	-3.53 (0.0030)	0.31 (0.0050)	0.49
11) White Males, 35-74 years, AHD3574	1316.76	-0.72 (0.0751)	-27.95 (0.0017)	1.98 (0.0140)	0.47
12) White Males, 45-64 years, AHD4564	597.13	-0.33 (0.1219)	-12.86 (0.0047)	0.98 (0.0203)	0.41

* The first number under TDS, P2AUTO and ATM7 are the beta values; the number in parentheses is the significance level of the F-statistic; the F-statistic describes the variables contribution to the model.

The TDS content of finished water could be increased by: 1) eliminating softening processes; 2) changing source of supply from surface water to ground water; or 3) adding soluble inorganic salts to water. The relationship of water softening and source of supply to CVD mortality rates was investigated; a statistically significant relationship was not found. Prior to implementation of any of the above procedures to increase the TDS of public water supplies, a more thorough understanding of the mechanism by which high total dissolved solids (if in fact TDS is the real 'water factor') acts to decrease CVD mortality would be both desirable and necessary.

The recent findings of Schroeder and Kraemer (32) that CVD mortality was inversely correlated with Langliers Index at a highly significant probability level can be used indirectly in the model to perform a benefit/cost analysis. Langlier's Index is a measure of the corrosiveness of water, a negative value is indicative of a corrosive water, a positive index indicates a non-corrosive water. Corrosive waters are usually softer and contain less dissolved solids than non-corrosive waters. Corrosive, or acidic water, can be stabilized by the addition of lime. Indirectly this can be viewed as the addition of dissolved solids to the water. It is estimated that adding 20 parts per million of CaO to the average water in this study would render it stable; this is equivalent to a 20 ppm increases in TDS. From equation 1, Table 11, the expected reduction in CVD mortality due to an increase in TDS is calculated to be 2.8 per 100,000. This is a decline in the mortality rate of approximately 0.5 percent, or decrease in total cardiovascular mortality in the 53 SMSA's of 1400.

Although it sometimes appears as though no expenditure is too great in protecting human health and life (consider the cost of maintaing persons on kidney dialysis equipment), most public health programs, due to the limited nature of public funds, must be examined within the benefit/cost framework. Ideally, the benefit/cost analysis should also consider alternative uses of public funds to ensure that they are invested in a manner such that the overall benefits are maximized.

In order to determine the benefits which would result from an improvement in water quality, i.e. a decrease in CVD due to an increase in TDS, decreases in both mortality and morbidity must be considered. The benefits resulting from a decrease in morbidity can be categorized as either direct or indirect. The direct benefits accrue due to decreased expenditures for hopital care, nursing and physicians' services, drugs and medical supplies, rehabilitation, construction, and similar items. The indirect benefits due to decreased morbidity result from a decline in lost productivity due to illness and disability. The expected benefits accruing as a result of reduced mortality from a change in the chemical composition of the water is a function of the value placed upon a human life; this is difficult to assess, particularly on a philosophical basis, however, on an economic basis, the value of a human life can be expressed as the present value of expected future earnings. Pyatt and Rogers (55)

suggest a methodology which can be used to evaluate a human life in these terms. Determination of the morbidity benefits is usually easier to assess than the mortality benefits.

According to Pyatt and Rogers (55), the most desirable way of expressing the costs of a municipal water supply system is in terms of total annual charges, including not only operating and maintenance costs, but also depreciation. Both costs and benefits should be expressed in constant dollars; that is in the same base year. The equation presented by Pyatt and Rogers (55) can be modified and used in the present benefit/cost analysis:

$$B/C = \frac{\sum_{n=0}^{a} \left[\frac{1}{(1+i^n)}\right] [Dn + In + Mn]}{C' + \sum_{n=0}^{a} \left[\frac{1}{(1+i)^n}\right] Cn}$$

Where, Dn is the direct morbidity benefit for the year n
In is the indirect morbidity benefit for the year n
Mn is the mortality benefit for the year n, expressed as a present value
Cn annual costs for year n
C' initial capital costs
$\frac{1}{(1+i)^n}$ is a discount factor which reduces costs or benefits to an equivalent investment in the base year

Table 12 shows the estimated initial investment and annual costs per capita, as well as the total annual costs for a corrosion control program. The capital equipment is assumed to have a life of ten years, and for the current analysis straight line depreciation is assumed to be adequate. The population and interest rate are considered to be constant at 51 million and 5 percent, respectively, for the ten year life of the equipment.

Table 12. Cost of Corrosion Control Program

	Per Capita	53 SMSA's (000,000)
Capital Equipment	$ 0.10	$ 5.10
Annual Depreciation	0.01	0.51
Annual Maintenance and Operating Costs	0.08	4.08

The economic cost of cardiovascular diseases and the benefits resulting from a decrease in CVD mortality for 1962 are shown in Table 13. As previously discussed, the general model predicts that a 20 ppm increase in TDS would result in a 0.5 percent decrease in the CVD mortality rate

TABLE 13. Estimated Economic Costs of Cardiovascular Diseases, by Type of Cost, 1962.*

Type of Cost	U.S.	53 SMSA's**	Benefits (53 SMSA's)
Morbidity Costs			
Hospital Care	$1234.8	$343.3	$1.72
Nursing Home Care	299.8	83.3	0.42
Physicians' Services	701.4	195.0	0.98
Drugs	279.4	77.7	0.39
Nursing Services	64.3	17.9	0.09
Other Health Services	34.4	9.6	0.05
Construction	190.0	52.8	0.26
Direct Costs	2804.1	779.6	3.91
Indirect Costs	2970.5	825.8	4. 4.13
Mortality Costs			
Present Value	15400.0***	4281.2	21.41
TOTAL COSTS	21174.6	5886.6	29.45

* From the Report to the President

** 28 percent of the United States

*** Estimated

(2.8 fewer CVD deaths per 100,000). The following assumptions are made in order to expedite evaluation of the expected benefits:

1. Morbidity will also be reduced by 0.5 percent.
2. All CVD related expenditures will decline by 0.5%percent.
3. The annual benefits (1962 dollars).
4. The discount rate (5%) will remain.
5. Since 28 percent of all CVD mortality in the U.S. in 1960 occurred in the 53 SMSA's, 28 percent of total expenditures is attributable to CVD in these SMSA's.

The cumulative benefits, costs and benefit/cost ratios, as shown in Table 14, were calculated from the equation presented earlier.

The present value of the benefits which would occur during the ten years is nearly $240 million; the present value of the expenditures necessary to realize these benefits during the same period is about $43 million, a B/C ratio of 5.6.

If the B/C ratio exceeds unity, it is usually recommended that the project be undertaken. It should be pointed out that if a favorable benefit/cost ratio is to be maintained, another capital investment will have to be made in the eleventh year. This results in a temporary decrease in the B/C ratio, although it will still be greater than one.

It is realized that the preceding benefit/cost analysis is somewhat crude, however, for the purposes of the current investigation it is felt to be adequate. If a more detailed analysis is desireable, it is recommended that the methodology outlined by Pyatt and Rogers (55) be followed.

Table 14. Cumulative Benefit/Cost/Ratios

Year	n	Costs (10^6)	Σ Costs	Benefits (10^6)	Σ Benefits	B/C
1962	0	9.69	9.69	29.45	29.45	3.04
1963	1	4.47	14.16	28.05	57.50	4.06
1964	2	4.25	18.41	26.72	84.22	4.57
1965	3	4.05	22.46	25.43	109.65	4.88
1966	4	3.77	26.23	24.22	133.87	5.10
1967	5	3.59	29.82	23.06	156.93	5.26
1968	6	3.42	33.24	21.96	178.89	5.38
1969	7	3.26	36.50	20.92	199.81	5.47
1970	8	3.11	39.61	19.93	219.74	5.55
1971	9	2.96	42.57	18.98	238.72	5.61

Although the independent variables in the general model developed as a result of this study have been formed to be significantly correlated with cardiovascular disease mortality by other investigators (14, 37, 44), it is still not possible to make a definitive statement concerning the true meaning of the relationship. Dudley, et.al., observed a high statistical

correlation between relative humidity and mortality (14). They suggested that this climatic variable may actually be an index of environmental stress. Examination of Table 6 indicates the relative humidity (RH7) in the present investigation was also correlated with CVD mortality at highly significant levels. The correlation coefficient between RH7 and ATM7 was 0.78; it was observed that if RH7 was substituted for ATM7 in the general model, the average S-squared (12 equations) did not change significantly, although it was less than the model containing ATM7. Perhaps the relative humidity in July is a etiologic factor which is being represented by the atmospheric pressure.

Or perhaps, altitude (ALT) is the fuel factor; the correlation coefficients between ALT and ATM7, and RH7 are -0.82 and -0.84, respectively. The correlation coefficients between altitude and the 12 dependent mortality variables were among the highest observed for any independent variable (see Table 6). Physiological investigations of people living at high altitudes have shown an inverse relationship between altitude and CVD mortality.

The percent of households owning two or more automobiles (P2AUTO) was found to be highly correlated with both the population density in the central city (CORDEN) and the percent of families living in one unit structures (P1UNIT), -0.56 and 0.65, respectively. These relationships are as expected, since the proportion of single unit structures increases and population density decreases as the distance from the city center increases, or in other words, as the suburban areas are approached. Ownership of more than one automobile becomes increasingly important in the absence of public transportation. Perhaps P2AUTO is representative of other factors, such as stress due to crowding or socioeconomic status. P2AUTO was found to be significantly correlated with education and employment, but not with income. Gradner, et.al. (44) considered the correlation among several socioeconomic variables and suggested that automobile ownership maybe an index of socioeconomic status.

With regard to TDS, the role of this variable in CVD mortality is also difficult to assess. Total dissolved solids were found to be significantly correlated with several other constituents of drinking water (K, Ca, Mg, hardness) all of which have been found to be statistically related to cardiovascular diseases, in numerous other studies. Which factor, if any, in water is etiologically related to CVD can not be specificied at the present time.

It is obvious from the literature reviewed and the results of this investigation that the apparent relationship between CVD mortality and environmental factors is difficult to ascertain, and also somewhat obscure. A reasonable conclusion based on the foregoing review and analysis, is that mortality from diseases of the cardiovascular system is a function of many factors, three of which may be related to the chemical composition of the water supply, the socioeconomic environment and the climate.

In general, it is felt that the statistical investigations reviewed here have reached a point of diminishing returns in using currently available data: a significant statistical relationship between some 'water factor' and CVD mortality is well established. Future investigations should entail the collection of original data by studying selected populations from various geographic and socioeconomic environments, and as suggested by Schroeder and Kraemer (32), the trace elements and chemical composition of tap water with respect to heart disease should be thoroughly investigated.

APPENDIX E
LIST OF ARTICLES AND REFERENCES
USED IN APPENDICES

1. Moriyama, I.M., D.F. Krueger and J. Stamler, "Cardiovascular Diseases in the United States," Harvard University Press, Cambridge, (1971).

2. World Health Organization, "Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death," World Health Organization, Geneva, 1957 (1955).

3. "A National Program to Conquer Heart Disease, Cancer and Stroke," Report to the President, The President's Commission on Heart Disease, Cancer and Stroke, Vol. II (1965).

4. Stamler, J., "Cardiovascular Diseases in the United States," Am. J. Card., 10:319-340 (1962).

5. National Institutes of Health, "Atherosclerosis: A Report by the National Heart and Lung Institute Task Force on Atherosclerosis," Vol. II, DHEW, Pub. No. (NIH) 72-219 (June 1971).

6. Epstein, F.H., "The Epidemiology of Coronary Heart Disease," J. Chron. Dis. 18:735-774,(1965).

7. Sauer, H.I., "Geographic Differences in the Risk of Premature Death," Business and Government Review, May-June, 19-26 (1970).

8. Enterline, P.E., A.E. Rikli, H.I. Sauer and M. Hyman, "Death Rates for Coronary Heart Disease in Metropolitan and Other Areas," Pub. Health Rep. 75:759-766 (1960).

9. Sauer, H.I. and P.E. Enterline, "Are Geographic Variations in Death Rates for Cardiovascular Diseases Real," J. Chron. Dis. 10:513-524 (1959).

10. Sauer, H.I., "Epidemiology of Cardiovascular Mortality--Geographic and Ethnic," A.J.P.H., 52:94-105 (1962).

11. Perry, H.M., Jr., "Hypertension and the Geochemical Environment," Ann. N.Y. Acad. Sci. 199:202-216 (1972).

12. Masironi, R., A.T. Miesch, M.D. Crawford and E.I. Hamilton, "Geochemical Environments, Trace Elements and Cardiovascular Diseases," Pan. Am. Hlth. Org. Bull. 1:53-64 (1973).

13. Shroeder, H.A. Relation Between Mortality from Cardiovascular Disease and Treated Water Supplies," J.A.M.A. 172:1902-1908 (1960).

14. Dudley, E.F., R.A. Beldin, B.C. Johnson, "Climate, Water Hardness and Coronary Heart Disease," J. Chron. Dis. 22:25-48 (1969).

15. Correa, P., J.P. Strong, "Atherosclerosis and the Geochemical Environment: A Critical Review," Ann. N.Y. Acad. Sci. 199:217-228 (1972).

16. Kobyashi, J., "On Geographical Relationship Between the Chemical Nature of River Water and Death Rate of Apoplexy," Ber Ohara Inst. Landw. Forsch. 11:12-21 (1957).

17. Schroeder, H.A. Deperative cardiovascular Disease in the Orient," J. Chron. Dis. 8:312-333 (1958).

18. Morris, J.N., M.D. Crawford and J.A. Heady, "Hardness of Local Water Supplies and Mortality from Cardiovascular Disease, Lancet 860-862 (April 22, 1961).

19. Muss,D.L. "Relationship Between Water Quality and Deaths from Cardiovascular Disease," J.A.W.W.A. 1371-1378 (November 1962).

20. Biorck, G, H. Bostrom and A. Widstrom, "On the Relationship Between Water Hardness and Death Rate of Cardiovascular Diseases," Acta Med. Scan. 178L239-252 (1965).

21. Bostrom, H. and P.O. Wester, "Trace Elements in Drinking Water and Death Rate in Cardiovascular Disease," Acta. Med. Scan. 181:465-473 (1967).

22. Anderson, T.W., W.H. LeRiche and H.S. MacKay, "Sudden Death and Ischemic Heart Disease Correlation with Hardness of Local Water Supply. New Eng. J. Med. 280:805-807 (1969).

23. Anderson T.W. andW.H. LeRiche, "Suddent Death from Ischemic Heart Disease in Ontario and Its Correlation with Water Hardness and Other Factors," C.M.A. Journal 105:155-160 (1971).

24. Peterson, D.R., D.J. Thompson and Jun-Mo Nam, "Water Hardness, Arteriosclerotic Heart Disease and Sudden Death," Am. J. Epidem. 92:90-3 (1970).

25. Neri, L.C., D. Hewitt and J.S. Mandel, "Risk of Sudden Death in Soft Water Areas," J. Epidem. 94:101-104 (1971).

26. Masironi, R., "Cardiovascular Mortality in Relation to Radioactivity and Hardness of Local Water Supplies in the U.S.A.," Bull. WHO 43: 687-697 (1970).

27. Robertson, J.S., "The Water Story," Lancet. 6:1160 (1969).

28. Croover, M.E. and J.E. Fulghum, "Death Rates Following a Sudden Change in Hardness of Drinking Water."

29. Crawford, M.D., M.J. Gardner and J.N. Morris, "Changes in Water Hardness and Local Death Rates," Lancet. ii:327-329 (1971).

30. Schroeder, H.A., "Municipal Drinking Water and Cardiovascular Death Rates," J.A.M.A. 195:81-85 (1966).

31. Sauer, H.I., "Relationship Between Trace Element Content of the Drinking Water and Chronic Diseases." Presented at the 16th Water Quality Conference, University of Illionis at Urbana, IL (1974).

32. Schroeder, H.A. and L.A. Kraemer, "Cardiovascular Mortality, Municipal Water and Corrosion," Arch. of Environ. Health. 28:303-311 (1974).

33. Crawford, M.D., M.J. Gardner and J.N. Morris, "Mortality and Hardness of Local Water-Supplies," Lancet. i:827-831 (1968).

34. Crawford, M.D., M.J. Gardner and J.N. Morris, "Cardiovascular Diseases and the Mineral Content of Drinking Water," Br. Med. Bull. 27: 21-24 (1971).

35. Comstock, G.W., "Fatal Arteriosclerotic Heart Disease, Water Hardness at Home and Socioeconomic Characteristics," Am. J. Epidem. 94:1 (1971).

36. Neri, L.C., J.S. Mandel and D. Hewitt, "Relation Between Mortality and Water Hardness in Canada, "Lancet. i:931-934 (1972).

37. Sauer, H.I., D.W. Parke and M.L. Neill, "Associations Between Drinking Water and Death Rates," Pam. Am. Health Org. Bull. 7:318-325 (1973).

38. Mulcahy, R., "The Influence of Water Hardness and Rainfall on the Incidence of Cardiovascular and Cerebrovascular Mortality in Ireland," J. Irish Med. Assoc. 55:17-18 (1964).

39. Mulcahy, R., "The Influence of Water Hardness and Rainfall on Cardiovascular and Cereb ovascular Mortality in Ireland." J. Irish Med. Assoc. 59:14 (1966).

40. Linderman, R.D., and J.R. Assenzo, "Correlations Between Water Hardness and Cardiovascular Deaths in Oklahoma Counties," Am. J. Pub. Health. 54:1071-1077 (1964).

41. Strong, J.P., P. Coreea and L.A. Solberg, "Water Hardness and Atherosclerosis, "Lab. Invest. 18:620-622 (1968).

42. Lehr, I., H.B. Messinger and R.H. Rosenman," A Sociobiological Approach to the Study of Coronary Heart Disease," J. Chron. Dis. 26: 13-30 (1973).

43. Masironi, R, "Trace Elements and Cardiovascular Diseases," Bull. W.H.O., 40:305-312 (1969).

44. Gradner, M.J., M.D. Crawford and J.N. Morris, "Patterns of Mortality in Middle and Early Old Age in the County Burroughs of England and Whales," Br. J. Prev. Soc. Med. 23:133 (1969).

45. Britt, F.W., D.G. Clayton, M.D. Crawford and J.N. Morris, "Clinical and Biochemical Indicators of Cardiovascular Disease Among Men Living in Hard and Soft Water Areas," Lancet i:122 (1973).

46. Bierenbaum, M.L., A.I. Fleischman, J.P. Dunn, H.D. Belk and E.M. Storter, "Serum Lipids in Hard and Soft Water Communities," Israel J. Med. Sci. 5:657-660 (1979).

47. Crawford, T. and M.D. Crawford, "Prevalence and Pathological Changes of Ischaemic Heart-Disease in a Hard-Water and in a Soft-Water Area."

48. Crawford, M.D. and T. Crawford, "Lead Content of Bones in a Soft and a Hard Water Area," Lancet. i:699-701 (1969).

49. Durfor, C.N. and E. Becker, "Public Water Supplies of the 100 Largest Cities in the United States," 1962 Geological Survey Water-Supply Paper 1812, U.S. Government Printing Office, Washington, D.C. (1964).

50. U.S. Weather Bureau, "Climatological Data," U.S. Department of Commerce, Washington, 46-48 (1959-1961).

51. Duffey, E.A. and R.E. Carroll, "United States Metropolitan Mortality, 1959-1961," P.H.S. Publication No. 999 Ap 39, National Center for Air Pollution Control, Cincinnati (1967).

52. Edwin Goldfield (Major Investigator), "County City Data Book, 1962," Bureau of Census, U.S. Department of Commerce.

53. "Census of Population, 1960," U.S. Bureau of Census.

54. Barr, A.J. and J.H. Goodnight, "Statistical Analysis System," North Carolina State University (1972).

55. Pyatt, E.E. and P.P. Rogers, "On Estimating Benefit-Cost Ratios for Water Supply Investments," Am. J. Pub. Health, 52:1729-1742 (1962).